(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2002 (10.05.2002)

PCT

(10) International Publication Number
WO 02/37856 A1

(51) International Patent Classification[7]: H04N 7/18

(21) International Application Number: PCT/US01/42912

(22) International Filing Date: 5 November 2001 (05.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/245,710 6 November 2000 (06.11.2000) US

(71) Applicant: DYNAPEL SYSTEMS, INC. [US/US]; 380 Lexington Avenue, Suite 4500, New York, NY 10168-1495 (US).

(72) Inventors: EDELSON, Steven, D.; 7 Sears Road, Wayland, MA 01778-2101 (US). DIEPOLD, Klaus; AM Stiergarten 10, 85635 Siegertsbrunn (DE).

(74) Agent: AITKEN, Richard, L.; Venable, Baetjer, Howard & Civiletti, LLP, 1201 New York Avenue, Suite 1000, P.O. Box 34385, Washington, DC 20043-9998 (US).

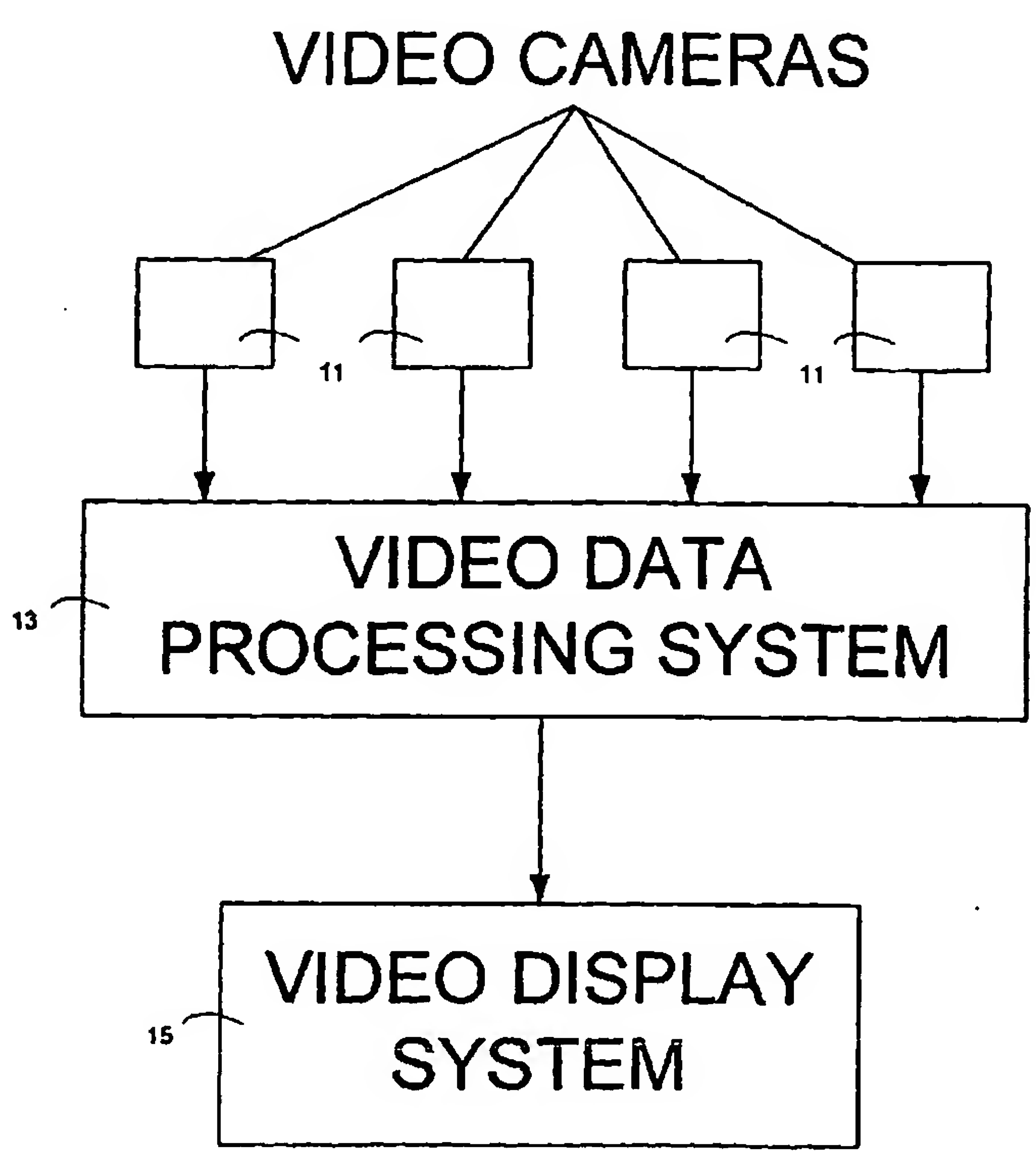
(81) Designated State *(national)*: JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SURVEILLANCE VIDEO CAMERA ENHANCEMENT SYSTEM



(57) Abstract: In a video camera surveillance system, a video processor determines dense motion vector fields between adjacent frames of the video. From the dense motion vector fields moving objects are detected and objects undergoing unexpected motion are highlighted in the display fo the video. To distinguish expected motion from unexpected motion, dense motion vector fields are stored representing expected motion and the vectors representing the moving object are compared with the stored vectors to determine whether the object motion is expected or unexpected. In an alternative embodiment, the video surveillance system comprises a panning camera and the frames of the video are arranged in a mosaic. Object motion in the video is detected by means of dense motion vector fields and the predicted position of objects in the mosaic is detected based on the detected object motion. The position of moving objects in the current frame being detected by the panning camera is compared with the predicted position of the objets in the mosaic and if the positions are substantially different, the corresponding object is tagged and highlighted as undergoing unexpected motion. A system is also disclosed for using the dense motion vector fields to control the motion of the panning camera to follow a moving object.

WO 02/37856 A1

Surveillance Video Camera Enhancement System

Background of the Invention

This invention relates to surveillance video camera systems and more particularly to surveillance video camera systems enhanced by detecting object motion in the video to reduce overload on the operator's attention.

In typical video camera surveillance systems of the prior art, multiple cameras are focused on multiple scenes and the resulting videos are transmitted to a monitoring area where the videos can be observed by the operator. The resulting multiple video motion pictures are simultaneously displayed or are displayed in sequence and it is difficult for the operator to detect when a problem has occurred in the detected scenes because of the large number of scenes which have to monitored. In some systems, a video camera is panned to increase the area that is monitored by a given camera. While such a system provides surveillance over a wide area, only part of the wide area is actually viewed at a given time leaving an obvious gap in the security provided by the scanning camera. To combat this latter problem, one system of the prior art combines the frames generated by the scanning camera into a mosaic so that entire scanned scene is displayed to the operator as an expanded panoramic view. In this system, each new video frame is compared with the previous detected frames displayed in the panoramic view and any differences are outlined thus providing an indication to the operator that the position of an object in the panoramic scene has changed. This system, while an improvement, nevertheless leads to an overload on the operator's attention, since all objects in this panoramic scene which undergo a change in position will be outlined and it is still difficult for the operator to recognize that one or more of the changes may represent a problem which requires attention. In addition, the fact that an object has undergone a change in position in many instances will not be brought to the operator's attention until the camera has completed a scanning cycle and then only if the object location of an which is undergoing a change in position appears in two different frames in a scanning cycle. Accordingly, there is a need for a video camera system which immediately brings to the operator's attention any significant or unexpected motion, which might represent a security problem requiring the operator's immediate attention.

Summary of the Invention

In accordance with the present invention a video camera surveillance system is provided with a video processor which has the capability of immediately detecting any object motion in a detected scene and more particularly detecting the occurrence of unexpected motion in a detected scene. In accordance with one embodiment, a plurality of surveillance cameras are provided which feed the videos to a video processing system wherein the videos are analyzed to determine dense motion vector fields representing motion between the frames of each video. From the dense motion vector fields, the motion of individual objects in the detected scenes can be determined and highlighted so that they are brought to the operator's attention. In accordance with the invention, the video processor stores dense motion vector fields representing expected motion in a scene and the dense motion vector field detected from the monitored scene is compared with the stored dense motion vector field representing expected motion to determine whether or not any unexpected motion has occurred. If an object is undergoing unexpected motion, this object will be highlighted in a display of the monitored scene.

In accordance with another embodiment of the invention, the surveillance system comprises a panning camera which pans a wide scene. The frames produced by the video camera are combined into a mosaic representing a panoramic view of the scene scanned by the camera. By means of dense motion vector fields, object motion in the scene being monitored is detected and, based on the detected motion of the objects, the future movement of the objects is predicted. In those portions of the scanned scene not currently being detected by the video camera, the position of the objects undergoing motion is updated in accordance with the predicted motion. Thus the moving objects in the panoramic scene will all be shown undergoing motion and changing position in accordance with the predicted motion. As each new frame of the scanned scene is detected by the video camera, the mosaic is updated with the new frame. If a given object undergoing motion in the current detected scene is substantially displaced from the predicted position when the current detected frame containing such object updates the panoramic scene, such object is tagged as undergoing unexpected motion and the object is highlighted.

In the system of the invention, the scanning speed is sufficiently slow that each part of the scene will be detected several times during each scan so that any objects undergoing motion will immediately detected. Any object undergoing exceptional motion such as moving at a high rate of speed or not corresponding to expected motion as represented by stored dense motion vector fields, may also be highlighted in the currently detected frames as shown in the displayed panoramic scene.

By only highlighting unexpected motion or exceptional motion, the system of the invention prevents overload on the operator's attention and only brings to the operator's attention those situations in the surveilled scene which require his immediate attention and action.

Brief Description of the Drawings

Figure 1 is a block diagram of a surveillance video camera system in accordance with one embodiment of the invention.

Figure 2 is a flow chart illustrating the video processing carried out by the video processing system in the embodiment of figure 1.

Figure 3 illustrates a display created by the system in figure 1 wherein a moving object may be highlighted by showing a telescopic enlarged view of an area around a moving object.

Figure 4 illustrates another display which may be provided by the surveillance system shown in figure 1.

Figure 5 is a block diagram of another embodiment of the present employing a scanning video camera.

Figure 6 is an illustration of a mosaic display created by the system of figure 5.

Figure 7 is a flow chart illustrating the process carried out by the video processor by the system in figure 5.

Description of Preferred Embodiments

In the system of the invention as shown in Figure 1 a plurality of video cameras 11 are each arranged to detect a video image of an area to be monitored by the video camera surveillance system. Each camera will send a sequence of

video frames showing the corresponding monitored area to a video data processing system 13. The video data processing system typically will comprise a video processor for each video camera but a high speed video processor could be employed to process the sequence of video frames received from each of the cameras simultaneously. The video data processing system detects object motion represented in the video received from each camera, highlights selected moving objects in the video, and transmits the resulting video to a video display system 15 in which the video from the four cameras are displayed. The video display system may be a separate monitors to display the videos simultaneously, or the videos may be all simultaneously displayed on the screen of the single monitor. Alternatively, the videos from the separate cameras may be displayed in sequence on one or more video monitors.

In preferred embodiments, the video processor will detect unexpected motion of objects in the videos and will highlight the objects undergoing this unexpected motion. Objects undergoing motion which is expected, may be highlighted in a different way than the way the unexpected motion is highlighted.

A flow chart of the process carried out by the video processing system 13 on a video received from one of the cameras in shown in Figure 2. As shown in this Figure, the video from one of the cameras is first processed to detect dense motion vector fields representing the motion of image elements in the received video. Image elements are pixel sized components of objects depicted in the video and a dense motion vector field comprises a vector for each image element indicating the motion of the corresponding image element. A dense motion vector field will be provided between each pair of adjacent frames in the video representing the motion in the video from frame to frame. The dense motion vector fields are preferably generated by the process disclosed in co-pending application Serial No. 09/593,521 entitled "System for the Estimation of Optical Flow", filed June 14, 2000 and invented by Sigfriend Wonneberger, Max Griessl and Markus Wittkop. This application is hereby incorporated by reference.

From the dense motion vector fields, the moving objects in the video are identified and are selectively highlighted. In a simplified version of the invention, all moving objects could be highlighted simply by changing a characteristic of all of the pixels in each video frame corresponding to a motion vector having a

substantial magnitude. This operation would highlight any moving object in the video, but would subject the operator's perception to overload since significant motion requiring the operators attention would be, in many cases, overwhelmed by detected motion which does not require the operators attention such as expected motion or trivial motion. This problem can be dealt with in a simplified version of the invention by storing in the video processor dense motion vector fields representing expected motion in the video. The dense motion vector fields generated from the current video are then compared with the dense motion vector fields representing the expected motion. The pixels corresponding to motion which is expected are then highlighted with one form of highlighting or not highlighted and the pixels corresponding to unexpected motion are highlighted with different form of highlighting.

In a preferred embodiment, the dense motion vector fields are analyzed to identify the pixels of individual moving objects. In the case of a moving object, the dense motion vector fields for the image elements of the object will all be similar. For example, if the object is moving linearly, the dense motion field vectors of the image elements of the object will all be parallel and of the same magnitude. If the object is rotating about a fixed axis, the dense motion vector field for the image elements of the object will be tangential around the center of the rotating object and will increase in magnitude from the center the edge of the moving object. If an object is not moving linearly, the dense motion vector field pattern for the object will be more complex, but nevertheless will fall into an easily recognized pattern. The video processor identifies sets of contiguous pixels which correspond to the dense motion field vectors representing a moving object. These pixels will then correspond to the image elements of the moving object.

In the preferred embodiment, the video processor will store the dense motion vector fields representing expected motion in the scene detected by the video camera, such as the motion of a fan, the motion of a rotisserie, or the motion of people walking along a walkway. When the detected object motion corresponds to the stored motion vectors representing expected motion, the video processor highlights the pixels of the object undergoing the expected motion in one selected way, such as tinting the pixels of the object undergoing expected motion blue. Alternatively, the pixels of the object undergoing expected motion

could be left unchanged and unhighlighted. When the detected object motion does not correspond to expected motion as represented by the stored dense motion vector fields, the object undergoing the unexpected motion is highlighted in a different way such as being tinted red or surrounded by a halo, or alternatively as being subjected to a telescopic effect. In producing the telescopic effect, the video processor defines a high resolution viewing bubble around the object undergoing the unexpected motion or around the area of the unexpected motion and magnifies it as shown in Figure 3. The operator may be given the ability to electronically steer the magnified viewing bubble around the scene to more clearly view items of interest. In a preferred embodiment, the unexpected motion is automatically highlighted by changing the pixel characteristic such as color or by adding a halo and then the operator can optionally define the high-resolution viewing bubble around the highlighted object after having his attention drawn to unexpected motion by the original highlighting.

In addition, in the preferred embodiment, the video processor can exclude from the highlighting process any trivial motion such as a motion of a small magnitude or a motion of a small object such as that of a small animal.

In the system as described above, the object having unexpected motion may be highlighted by changing its color, by changing its saturation, by changing it contrast, or by placing a halo around the object. Alternatively, the object may be highlighted by defocusing the background which is not undergoing unexpected motion or by changing the background to a grey scale depiction.

Another feature of the present invention is illustrated in Figure 4. In accordance with this feature of the invention, a moving object in the display is identified as described above by flagging the contiguous pixels representing the moving object. The velocity of the moving object is then detected from the dense motion vector field vectors representing the motion of the picture elements corresponding to the moving object. Information is then added to the display to indicate the speed and direction of the moving object as shown in Figure 4. The information may be in the form of an arrow indicating the direction of the motion and containing a legend in the arrow indicating the speed and feet per second and the heading of the object in degrees. In Figure 4, the cart being pushed by a customer is moving at 2.6 feet per second at a heading of 45°.

In accordance with a further feature of the invention, the position of the flagged moving object at a predetermined time in the future is predicted and the position of the moving object at this future time is then indicated in the display by a graphic representation such as showing a representation of the moving object in outline form.

Additional statistics may also be included in the display such as a time that the object has been shown in the display, the time duration from flagging of the object as a moving object, or other information related to the object motion.

In the embodiment of the invention shown in Figure 5, a panning camera 21 senses a wide scene by oscillating back and forth to scan the scene. The video produced by the camera is sent to a video processor 23, which arranges the received frames in a mosaic presenting a panoramic view of the scene scanned by the panning camera 21. The mosaic is transmitted to the video display device 25 where the mosaic of the scanned scene is displayed as shown in Figure 6 so that the viewer can view the entire scene scanned by the camera. As shown in Figure 6, the display will outline the currently received frame so that the viewer will have the information as to which part of the scanned scene is currently being received by the video camera.

The video processor, in addition to combining the received frames into a mosaic, detects object motion in the scanned scene and from the object motion detects the predicted position of any moving objects in the portions of the scanned scene not currently being detected. The video processor modifies the display of the moving objects in the portion of the scanned scene which are outside of the frame currently being detected by the camera to show the moving objects in their predicted positions in this portion of the scene being scanned. Then when the scanning camera returns to a portion of the scene containing the moving object shown in a predicted position, the position of the moving object will be undated in accordance with currently the detected frame containing a the moving object. In this way, the scene observed by the operator in the entire mosaic will show all the moving objects in their expected positions based on their detected motion.

When the actual position of a moving object is detected by the scanning camera and the object is substantially displaced from its predicted position, the object is tagged as having unexpected motion and the object is highlighted such as

by changing its color, by changing its brightness or saturation, by placing a halo around the object, or by magnifying the area around the object to provide a telescopic effect at the location of the object.

The camera is panned to scan the scene at a slow enough rate that so each location, is detected in several sequential frames during the scan of the camera. This feature enables the system of the invention, making use of dense motion vector fields to detect object motion, to detect any object motion during each scan of the scene.

The system of figure 5 may also detect unexpected motion by storing dense motions vector fields representing expected motion in the manner described above in connection with the embodiment of figure 1. Because the system of figure 5 detects object motion immediately, this form of unexpected motion may be immediately highlighted without waiting for the camera to again cycle through the same portion of the scene.

As a result of viewing the entire scanned scene, including predicted object motion in the scene, the operator may wish to get an immediate update of a specific object in the scanned scene. Rather that wait for the panning camera to again reach the object, the operator can cause the camera to snap to that view by means of servomechanism 27 for a real time display of the object of interest and can cause the camera to zoom in on the object if desired.

Figure 7 is a flow chart illustrating the operation of the video processor to make a mosaic of the received picture frames to display the entire scene scanned by the camera and to detect moving objects and to predict and display their predicted positions in the scanned scene. As shown in Figure 5 the video is first processed in step 31 to detect the dense motion vector fields representing the motion of image elements between the currently detected frame and the adjacent frames in the video. Since the camera is being panned, the dense motion vector field will represent the apparent motion of the background due to the camera motion as well as motion of objects relative to the scene background. From the dense motion vector fields, the camera motion is detected and the motion of objects, separated from the camera motion, is also detected in step 32. To detect the camera motion from the dense motion vector fields, the predominant motion represented by the vectors is detected. If most of the vectors are parallel and of

the same magnitude, this will indicate that the camera being moved in a panning motion in the opposite direction to that of the parallel vectors and the rate of panning of the camera will be represented by the magnitude of the parallel vectors. To detect the object motion, vectors corresponding to the camera motion are subtracted from the dense motion vector field vectors detected in the first instance between the adjacent frames. The resulting difference vectors will represent object motion. From the vectors representing object motion, the moving objects in the current frame are identified and their motion is determined. In step 33, the position of the currently detected frame in the mosaic is roughly determined from the detected camera motion. The currently detected frame may then be finely aligned with the mosaic by comparing the pixels at the boundary of the detected frame with the corresponding pixels in the same location in the mosaic. In step 34, the position of the moving objects in the current frame is compared with the predicted positions for these objects in the current frame. As will be explained below, the objects will be displayed in the mosaic in their predicted positions based on their previously detected motion. If the position of an object in the currently detected frame is not approximately the same as its predicted position in the mosaic, the object is tagged as having unexpected motion. In step 35, the mosaic is updated with the recurrent frame by replacing the pixels in the mosaic with the corresponding pixels of the current frame. At this time, the objects tagged in the current frame is undergoing unexpected motion are highlighted. In step 36, the position of all moving objects outside the currently detected frame are undated in accordance with there predicted positions. In this process, the objects which were previously flagged as being moving objects and which are outside of the currently detected frame have their current positions predicted based on the motion determined for the moving objects. To update the position of a moving object, the flagged pixels of the moving object replace the pixels in the mosaic at the predicted position of the moving object. The pixels of the moving object which are not replaced in this process (in the object's previous position) are replaced with corresponding background pixels in the scanned scene. The process then returns to step 31 to determine the dense motion vector field between the next detected video frame and the adjacent video frames and the process then repeats for the next received video frame from the panning camera.

As described above, objects undergoing unexpected motion are highlighted. In addition, any objects undergoing expected motion or undergoing substantial expected motion may be highlighted in a different manner to distinguish them from objects undergoing unexpected motion as described above in connection with the embodiment of Figure 1.

As described above, the panning camera may be zoomed in and out. While the camera is being zoomed in and out, the action of the camera is considered camera motion and the video frames produced during the zooming or while the camera is in a zoomed in or out state, can be added to the mosaic. In this process the zooming camera motion is detected by the prevailing motion vectors extending radially inwardly or outwardly. Once the camera motion has been detected, the size of the camera frames are adjusted to correspond to that of the mosaic frames and the currently detected frames are then located in the mosaic in the same manner as described above in connection with locating the camera frames produced by the camera panning motion.

In accordance with another feature of the invention, the video processor, controls the operation of servomechanism 27 to cause the panning motion of the camera to follow a moving object and keep the moving object centered in the detected frame. To carry out this control, the video processor determines the predicted immediate future locations of the moving objects. The predicted immediate future locations of the moving object are determined from the dense motion vector field vectors for the moving objects as explained above. By continuously moving the camera to the predicted immediate future locations of the moving object, the camera is made to follow the moving object keeping it centered in the currently detected frame.

In the above described systems, the location of where the videos are displayed may be at a position a long distance from the position of the surveillance cameras. In such an instance, to permit the data to be transmitted over the long distance by telephone line or by the internet, the transmitted data is compressed. In accordance with one embodiment, the video data is processed by a video processor at the location of the surveillance camera or cameras to identify and tag moving objects. Then after video has been transmitted to the display device representing the background being televised by the surveillance camera,

subsequent transmissions will only transmit the pixels representing the objects undergoing motion. This compression can be used with either of the embodiments described above.

Alternatively, the successive video frames transmitted to the receiver can be compressed by eliminating selected frames on the cameras side and then recreating these frames on the receivers side as described in co-pending application serial no. 09/816,117, filed February 26, 2001, entitled "Video Reduction by Selected Frame Elimination" or alternatively in application serial no. 60/312,063, filed August 15, 2001, entitled "Lossless Compression of Digital Video." These two co-pending applications are hereby incorporated by reference.

The Surveillance Video Camera Systems described above solve the problem of operator overload when a large amount of space has to be monitored by video cameras and makes it possible for the operator to detect and focus on important or unexpected motion when such motion occurs in the scene being monitored by the surveillance cameras.

The above description is a preferred embodiments of the invention and modifications may be made thereto without departed from the spirit and scope of the invention, which is defined in the appendant claims.

WHAT IS CLAIMED:

1. A surveillance method comprising detecting a scene to be surveilled with a video camera, processing the resulting video to detect moving objects in the detected scene, distinguishing objects undergoing unexpected motion from objects undergoing expected motion and highlighting the objects undergoing expected motion.

2. A surveillance method as recited in claims 1 further comprising storing representations of motion expected in the detected scene and comparing the motion of moving objects in the detected scene with the stored representations of expected motion to distinguish unexpected motion from expected motion.

3. A method as recited in claim 1 further comprising:
predicting the future positions of the detected moving objects;
comparing the actual positions of the moving objects with the predicted positions; and
identifying an object as undergoing unexpected motion when the positions of such object is substantially different than the predicted position for such object.

4. A method as recited in claim 1 wherein the objects undergoing unexpected motion are highlighted by magnifying the object and the area around the object identified as undergoing unexpected motion.

5. A surveillance method comprising detecting a scene to be surveilled with a video camera, detecting dense motion vector fields representing the motion of image elements from frame to frame in the video produced by said video camera, identifying moving objects depicted in said video by means of said dense motion vector fields and displaying said video with the identified moving objects highlighted in the display of said video.

6. A method as recited in claim 5 wherein a detected moving object is highlighted by magnifying the display of the moving object and the area around the moving object.

7. A surveillance method comprised of panning a video camera to detect a scene, combining the frames of the resulting video into a mosaic representing a panoramic view of said scene, detecting the motion of moving objects in said scene, determining the predicted position of objects in said scene determined from the detected motion of said objects, and updating the position of said moving objects in accordance with the predicted motion of said moving objects in said mosaic.

8. A surveillance method as recited in claim 7 further comprising comparing the position of detected moving objects in the current video frame detected by said video camera with the predicted position for said moving objects and flagging said moving objects as having said unexpected motion when the position of said objects in the current frame is substantially different than the predicted position for such objects.

9. A surveillance method as recited in claim 8 further comprising highlighting in said mosaic the objects flagged as having unexpected said motion.

10. A method of making a video of a moving object comprising detecting a scene containing said moving object, with a video camera to produce a video depicting said moving object, generating dense motion vector fields representing the motion of image elements from frame to frame in said video, determining the motion of said moving object from said dense motion vector field, predicting the immediate future position of said moving object from the detected motion of said moving object, and controlling the motion of said video camera in accordance with the predicted immediate future position of said moving object to maintain the moving object centered in the frame currently being detected by said video camera.

11. A surveillance system comprising a video camera arranged to detect a scene to be survielled, a video processor operable to detect moving objects in the video produced by said video camera and to distinguish objects undergoing unexpected motion from objects undergoing expected motion and to highlight the objects undergoing unexpected motion, and a display device connected to receive the processed video from the video processor to display the processed video with the objects undergoing unexpected motion highlighted.

12. A surveillance system as recited in claim 11 wherein said video processor is provided with a storage capacity which stores representations of expected motion in a detected scene, said video processor comparing the motion of moving objects in a detected scene with the stored representations of expected motion to distinguish unexpected motion from expected motion.

13. A surveillance system as recited in claim 11 wherein said video processor is operable to predict the future positions of the detected moving objects from the motion of the detected moving objects, to compare the actual positions of the moving objects with the predicted positions, and to identify an object as undergoing unexpected motion when the actual position of such object is substantially different than the predicted position for such object.

14. A surveillance system as recited in claim 11 wherein said video processor highlights an object undergoing unexpected motion by magnifying such object and the area around such object identified as undergoing unexpected motion.

15. A surveillance system comprising a video camera arranged to detect a scene to be surveilled to produce a video of said scene, a video processor connected to receive said video and operable to detect dense motion vector fields representing the motion of image elements from frame to frame in said video and to identify moving objects depicted on said video by means of said dense motion vector fields, and to highlight in said video a detected moving object by magnifying the display of the moving object and the area around moving object,

and a video display device connected to receive the processed video from said video processor and to display the processed video with the moving object and the area around the moving object being magnified.

16. A surveillance comprising a panning video camera arranged to detect a surveilled scene, a video processor connected to receive the video produced by said video camera and operable to combine the frames of said video into a mosaic representing a panoramic view of said scene, to detect the motion of moving objects in said scene, to determine the predicted position of objects in said scene determined from the detected motion of said objects and to update the position of said moving objects in said video in accordance with the predicted motion of said objects in said mosaic, and a video display device connected to receive the video processed by said video processor and to display said mosaic as a panoramic view of the surveilled scene with the moving objects depicted in their predicted positions.

17. A surveillance system as recited in claim 16, wherein said video processor is further operable to compare the positions of detected moving objects in the frame currently detected by said video camera with the predicted positions of said moving objects, to flag a moving object as having said unexpected motion when the position of such moving object in the current frame is substantially different than the predicted position for such object, and to highlight in said mosaic the object flagged as having unexpected motion whereby the display device displays said mosaic with the object undergoing unexpected motion highlighted.

18. A video system comprising a panning video camera a video processor connected to receive the video produced by said panning video camera and operable to generate dense motion vector fields representing the motion of image elements from frame to frame in the video produced by said video camera, to determine the motion of any depicted moving object in said video from said dense motion vector fields, and to predict the immediate future position of said moving object from the detected motion of said moving object, a controller for

controlling the motion of said video camera, said video processor controlling said controller to control the motion of said video camera in accordance with the predicted immediate future position of said moving object to maintain the moving object centered in the frame currently being detected by said video camera.

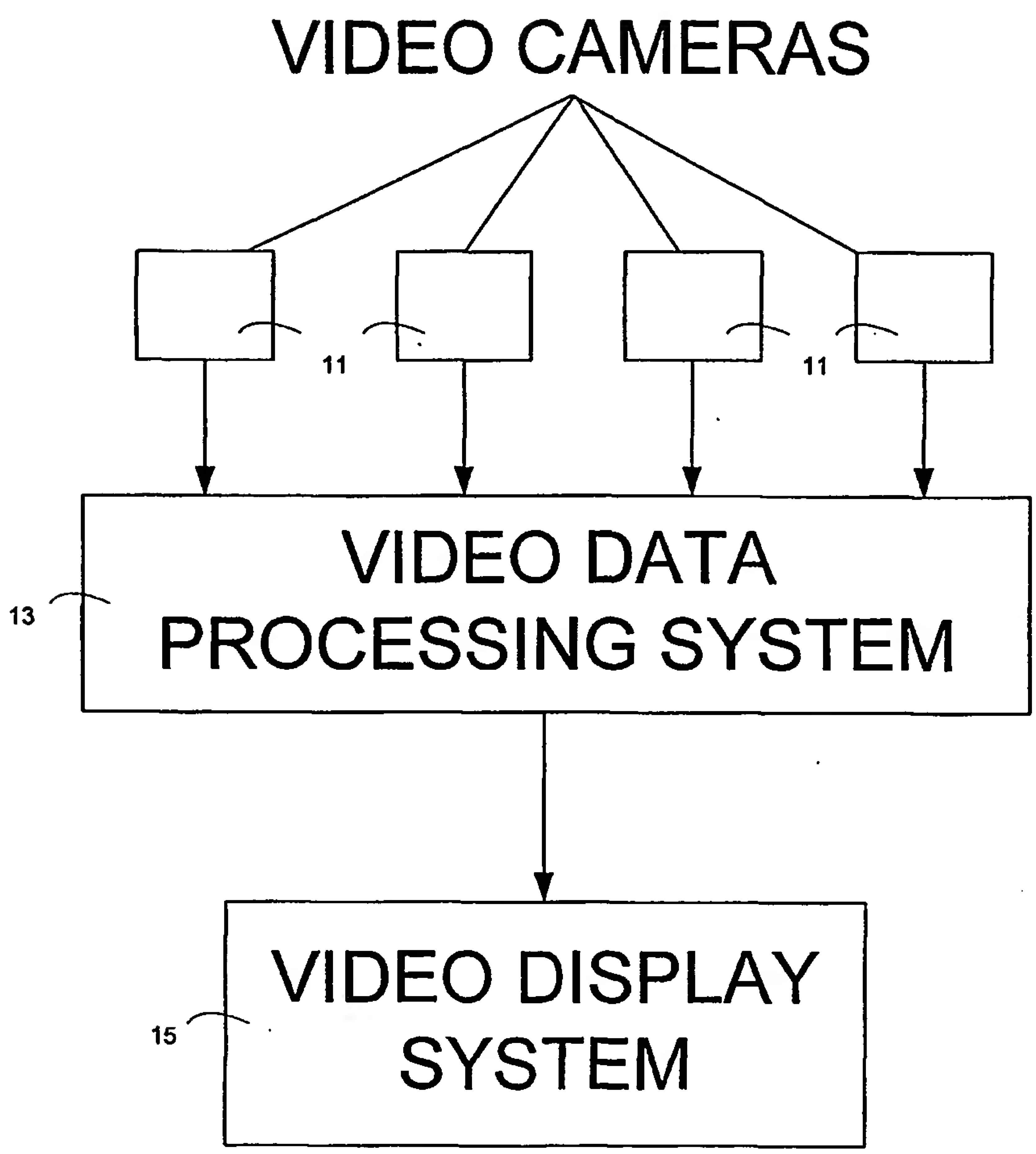

VIDEO CAMERAS
11
11
VIDEO DATA PROCESSING SYSTEM
13
VIDEO DISPLAY SYSTEM
15

Fig. 1

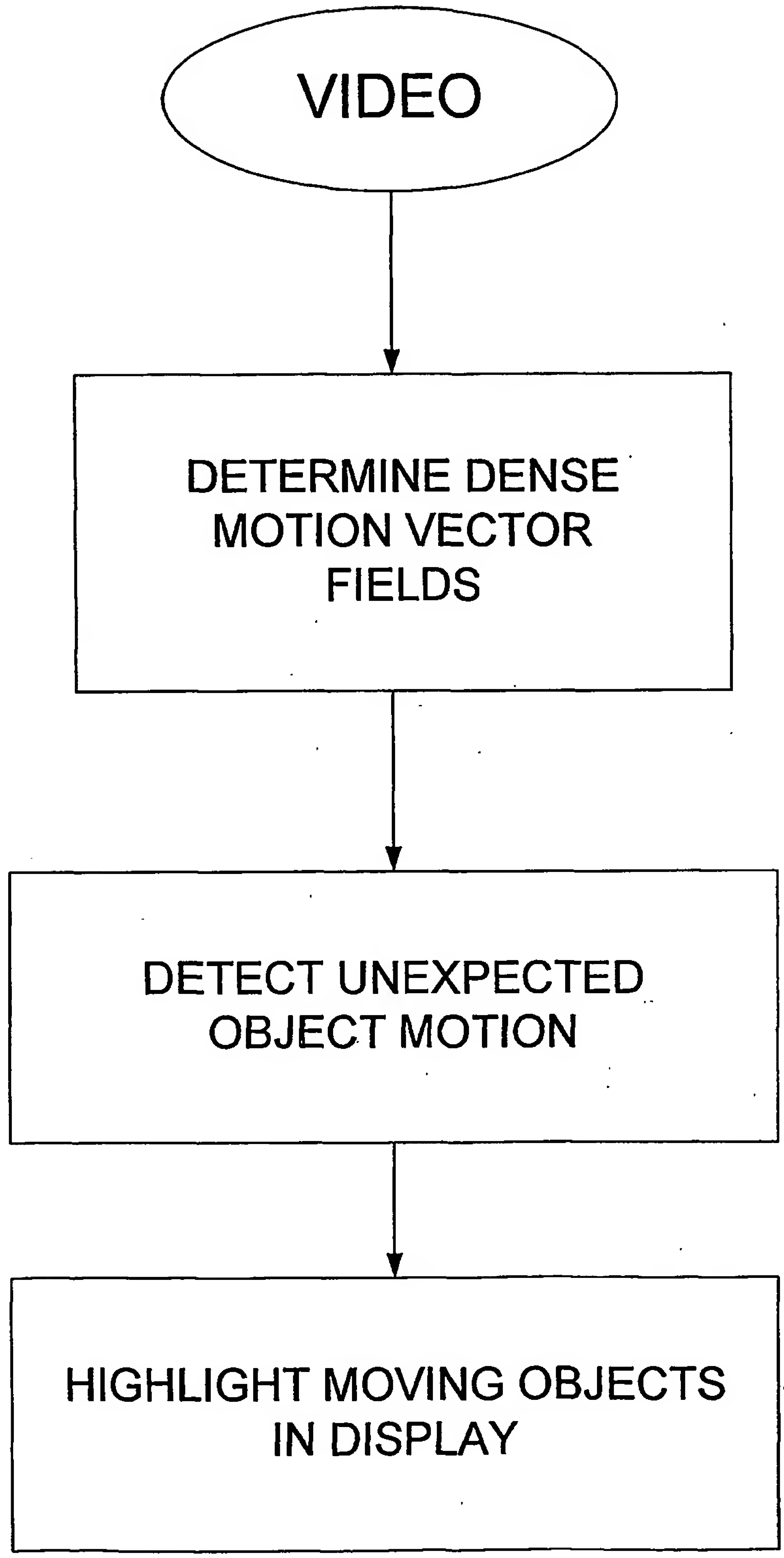

VIDEO
DETERMINE DENSE MOTION VECTOR FIELDS
DETECT UNEXPECTED OBJECT MOTION
HIGHLIGHT MOVING OBJECTS IN DISPLAY

Fig. 2

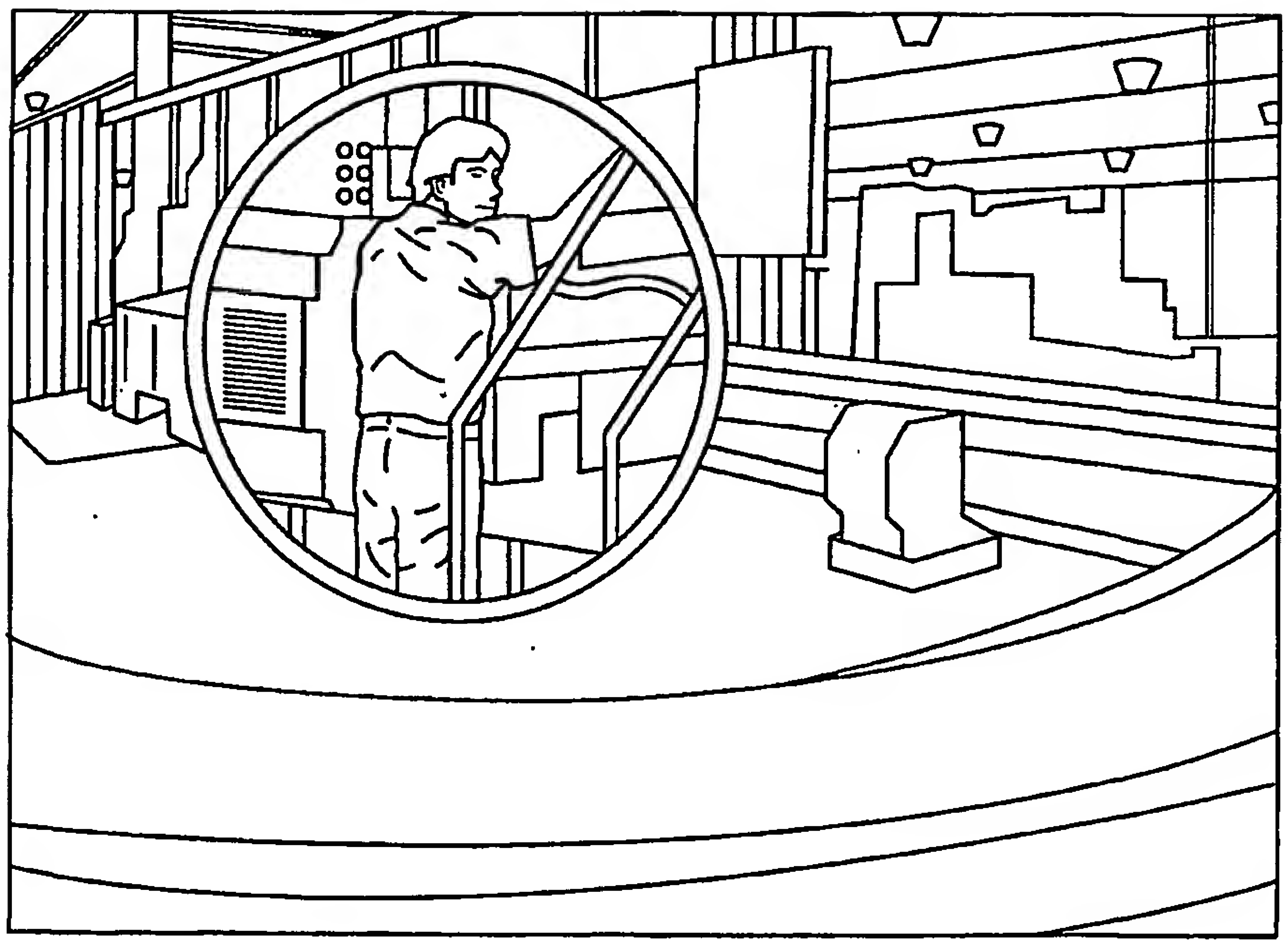


FIG. 3

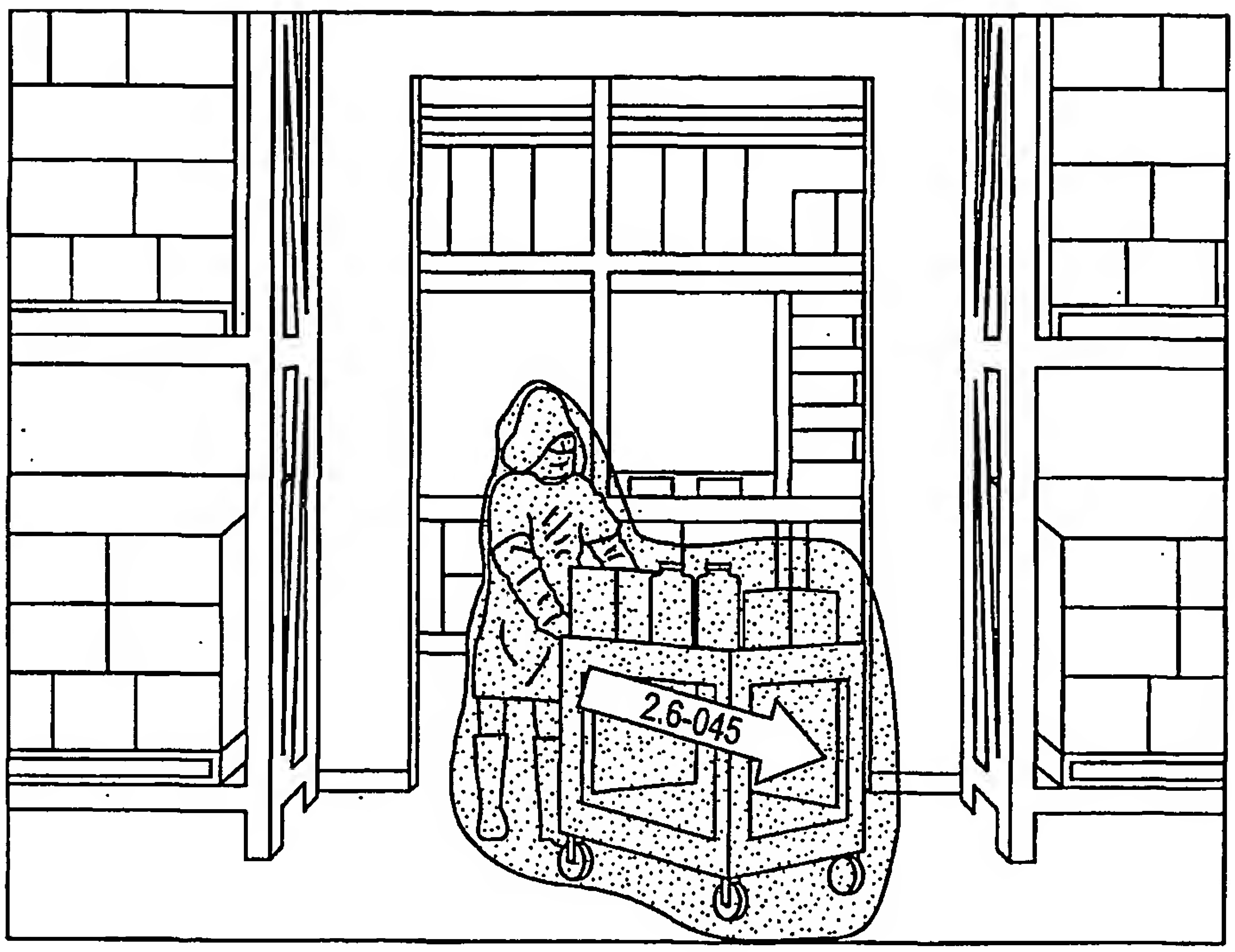


FIG. 4

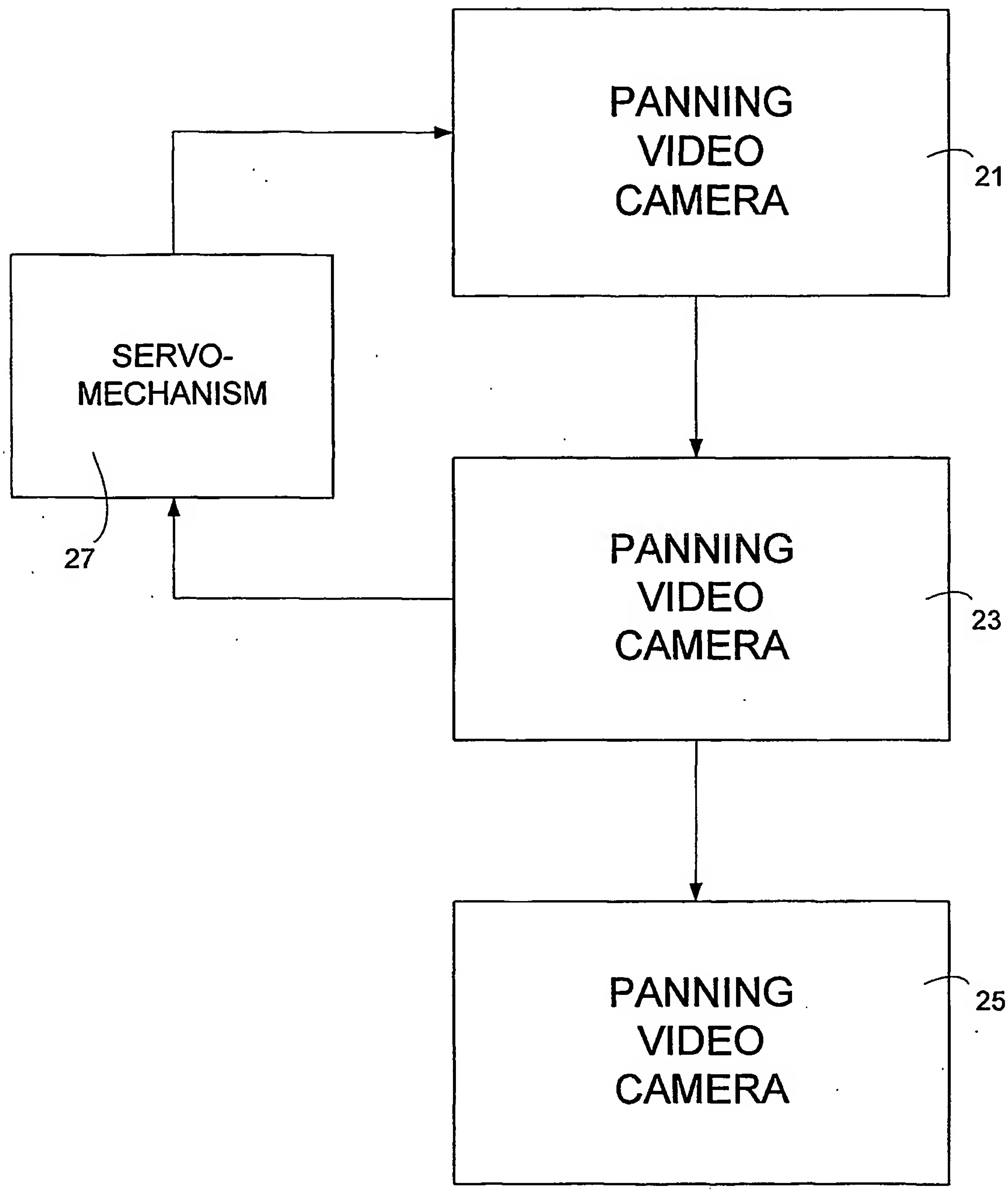

PANNING
VIDEO
CAMERA
21
SERVO-
MECHANISM
27
PANNING
VIDEO
CAMERA
23
PANNING
VIDEO
CAMERA
25

Fig. 5

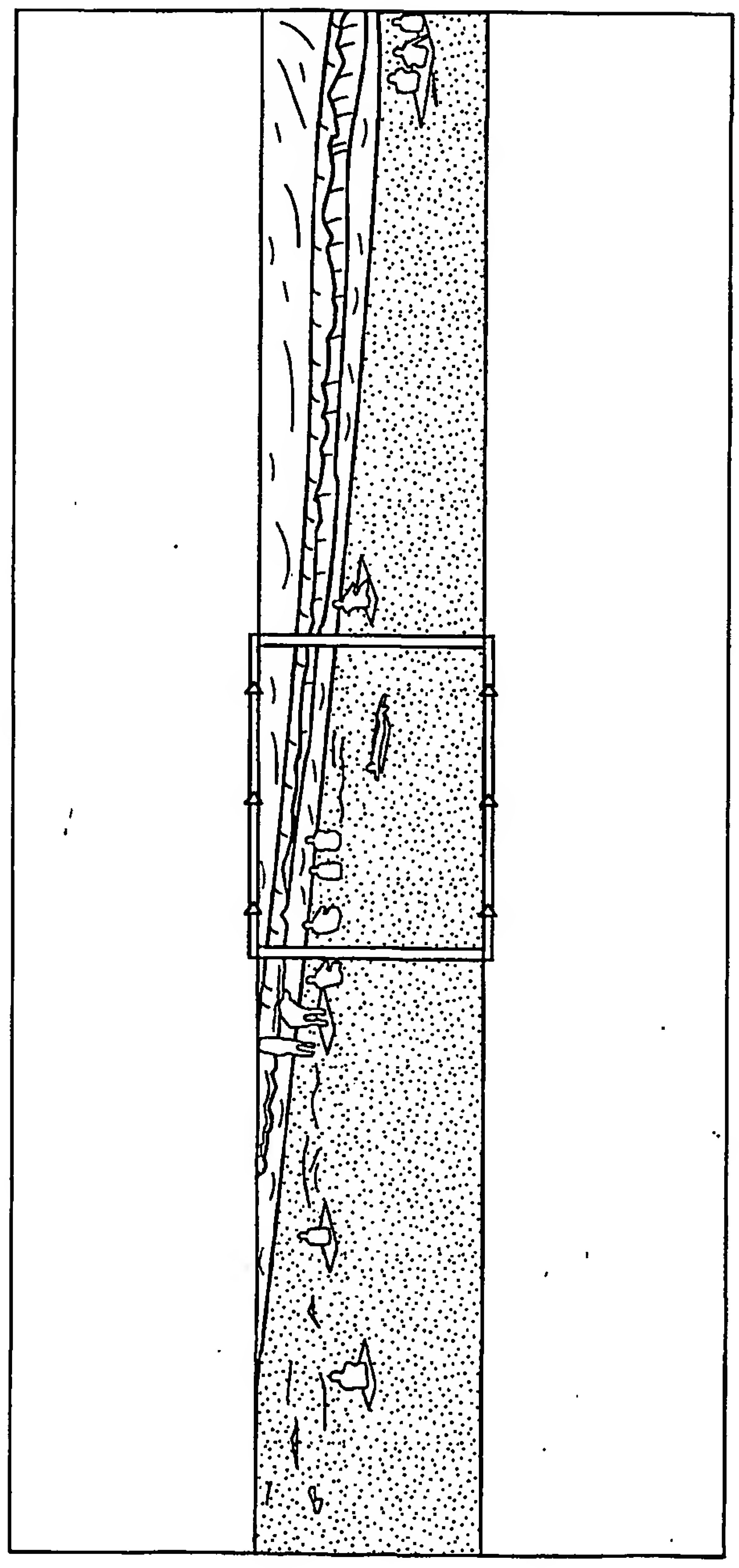


FIG. 6

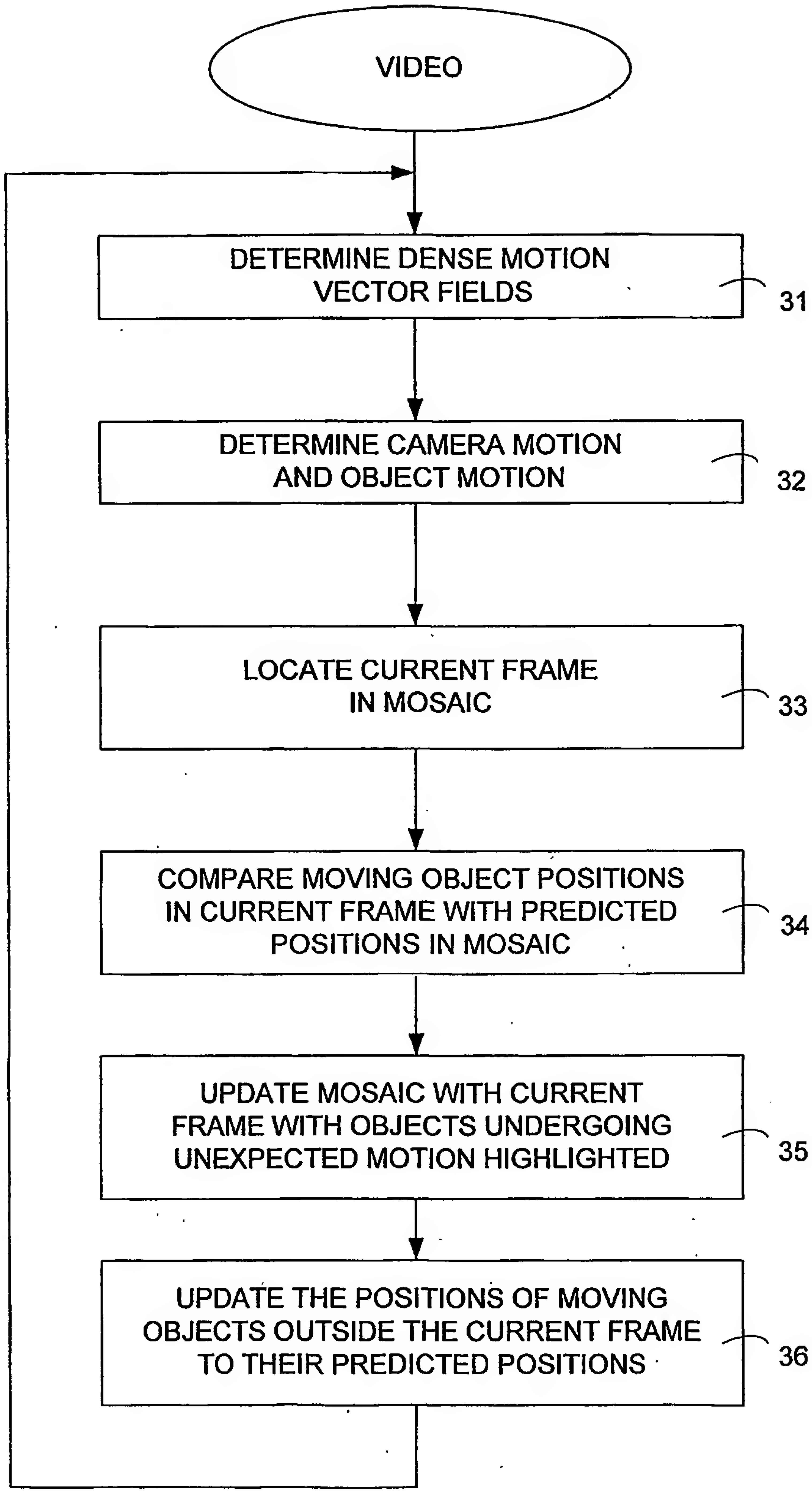

VIDEO
DETERMINE DENSE MOTION VECTOR FIELDS
31
DETERMINE CAMERA MOTION AND OBJECT MOTION
32
LOCATE CURRENT FRAME IN MOSAIC
33
COMPARE MOVING OBJECT POSITIONS IN CURRENT FRAME WITH PREDICTED POSITIONS IN MOSAIC
34
UPDATE MOSAIC WITH CURRENT FRAME WITH OBJECTS UNDERGOING UNEXPECTED MOTION HIGHLIGHTED
35
UPDATE THE POSITIONS OF MOVING OBJECTS OUTSIDE THE CURRENT FRAME TO THEIR PREDICTED POSITIONS
36

Fig. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/42912

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :H04N 7/18
US CL : 348/152
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,081,606A (HANSEN et al) 27 JUNE 2000, COL. 3 Lines 1-20	1-2,5
Y		3,4,6,12-14
Y	US 6,028,626A (AVIV) 22 FEBRUARY 2000, COL 5 LINES 55-68	3,4,6,12-14
A	US 6,091,771A (SEELEY et al) 18 JULY 2000, All	7,8,9,11-18
A	US 6,111,913A (MURDOCK et al) 29 AUGUST 2000, All	10

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 MARCH 2002	16 APR 2002

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	CHARLES PARSONS
Facsimile No. (703) 305-3230	Telephone No. (703) 305-3862

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/42912

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

348/152,154,155,143,213
375/240
382/107

52

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 February 2003 (13.02.2003)

PCT

(10) International Publication Number
WO 03/013113 A2

(51) International Patent Classification[7]: H04M 3/00

(21) International Application Number: PCT/GB02/03532

(22) International Filing Date: 31 July 2002 (31.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0118921.6 2 August 2001 (02.08.2001) GB

(71) Applicant (*for all designated States except US*): **EYRETEL PLC** [GB/GB]; Kings Court, Kingston Road, Leatherhead, Surrey KT22 7SZ (GB).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **BLAIR, Christopher, Douglas** [GB/GB]; Ivor Cottages, Balneath Farm, South Chailey, East Sussex BN8 4AP (GB). **PEREIRA, Ainsley, Malcolm** [GB/GB]; 4 Aspen Way, Horsham, West Sussex RH12 4AH (GB). **SAARISTE, Samuel** [NL/GB]; 12 Oaklands, Horsham, West Sussex RH13 5LJ (GB).

(74) Agent: **GILL, David, Alan**; W.P. Thompson & Co., Celcon House, 289-293 High Holborn, London WC1V 7HU (GB).

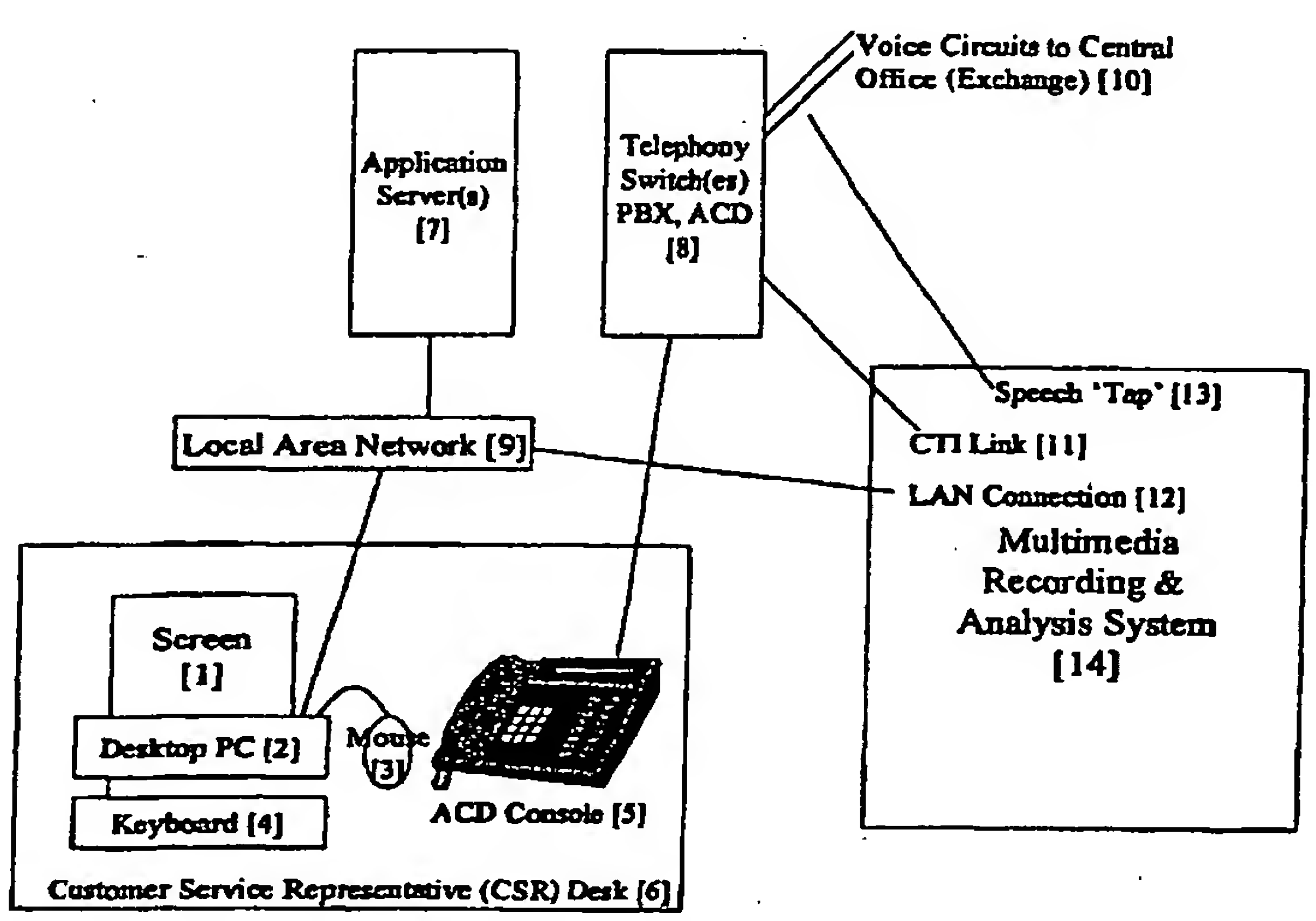
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: TELECOMMUNICATION INTERACTION ANALYSIS



(57) Abstract: The present invention provides for a method of monitoring sets of related communication signal streams comprising the steps of analysing the content or parameters associated with a component of one of the signal streams according to a first analysis criteria, analysing a second component of a related signal stream or parameter associated therewith, according to a second analysis criteria, providing results of the analysis of the said one of the signal streams and which is responsive to the said analysis according to the second criteria. Also, the analysis of the energy envelope representative of at least one communication signal can be provided for and the method of the present invention can further include steps of conducting speech recognition of the identification of words and/or phrases within a communications traffic stream and in which the scale and/or nature of recognition analysis applied to the speech recognition is varied responsive to the analysis of content or parameters associated with the communication stream.

WO 03/013113 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Telecommunication Interaction Analysis

The present invention relates to the analysis of communication signals and in particular such signals representing interaction between the users of a telecommunication system.

Commercial organizations have, for some time, taken the step of recording communications streams such as telephone calls between their staff and their customers. Traditionally this has been necessary to help satisfy regulatory requirements or to help resolve disputes. More recently, the emphasis has moved towards the review of such communications interactions from a quality perspective: the aim being to identify good and bad aspects and characteristics of communication exchanges with a view to improving the level of customer service given.

Also, a record of activity as occurring on an associated display such as a PC screen can also be made and can serve to improve the completeness of a communication-exchange review procedure. In this manner, the reviewer is then able to ascertain how accurately staff are entering information provided during a telephone conversation. Also, particular aspects of an employee's data entry skills and familiarity with the application can be reviewed by recording keystrokes and mouse movement/clicks etc.

So-called Call Detail Recording systems have been employed in order to allow for the prevention of abuse of telephone systems and to apportion costs to the relevant department or individual making the calls. Originally, such records were printed out directly from the Private Automatic Branch Exchange (PABX) onto a line printer. Systems are also now available that are able to store this information in a

database allowing more sophisticated reporting and for the searching of calls on the basis of one or more of the details related to the stored call.

Several systems have been developed, for example the AutoQuality system available from e-Talk, and the eQuality system available from Witness Systems and also the present applicant's QualityCall system, that employ call recording in combination with call detail recording and a database application to perform routine monitoring of calls with the intention of identifying weaknesses in the performance of individual Customer Service Representatives (CSRs). Typically a small percentage of the CSRs' calls are reviewed and scored against a set of predetermined criteria to give an indication of the performance of the member of staff.

Also of relevance is the current state of the art of speech recognition systems. First, the automation of simple interactions previously conducted via human interaction, or via touch tone menus, can be achieved. Secondly, dictation products are available that can translate the contents of an audio input into text even though they may exhibit error rates that are greater than would be acceptable if a meaningful transcription of the call was required.

Recording systems are also available that can be arranged to provide for the analysis of the content of, for example, a communications stream. Systems providing for the recording of particular events, or incidents, that might arise during a telephone conversation, and the time at which such events or incidents occur within a communications interaction have also been developed.

Such known systems however, and in particular quality-monitoring systems, exhibit disadvantages and limitations

and are discussed in International Application WO 01/52510.

For example, such systems tend to be extremely labour intensive. The time required to review an interaction can typically take at least as long as the original interaction lasted. It can also prove necessary to listen to and review the recording of the interaction several times. For example, an initial review may be required in order to determine the content and type of call, and whether or not it is complete enough and appropriate to allow for full evaluation. If so, it is then re-played completely for review against pre-determined scoring criteria. It then has to be re-played again for review with the CSR who took the call.

Known systems also prove unable to identify infrequent problems. Because of the time taken to review a call, it is rare that more than a fraction of one percent of all calls are evaluated and reviewed. This renders the reviewed calls statistically very poor for identifying rare problems. Realistically, such systems can only hope to provide an indication of the average quality of interactions carried out by each CSR.

Increasingly, CSRs are expected to be multi-skilled and to handle a wide range of different types of calls. Unless many more hours are spent reviewing calls, it is impossible effectively to identify problems that occur in a small proportion of a CSR's calls. If problems are only rarely spotted, it then becomes very difficult to recognize underlying patterns since such instances become isolated.

Also, such known systems are very much subjective and, even with the best training and call-evaluation coaching, the evaluator will apply at least some degree of subjectivity to their evaluation particularly with softer aspects of

assessment such as customer satisfaction levels. While such systems can provide tools that highlight discrepancies between different evaluators, they are restricted in that they cannot serve to prevent such subjectivity.

Known systems also are not generally normalized. For example, the manner in which organizations choose to measure call quality is entirely at their discretion and so a 95% quality rating achieved by one organization may in reality be worse than the 90% rating achieved by another organization employing a stricter marking schema. This lack of consistency between organizations makes it difficult, for example, for organizations to evaluate how they compare with their industry peers or indeed with other industries.

The present invention seeks to provide for a system and related method which can offer advantages over known such systems and methods.

According to an aspect of the present invention there is provided a method of monitoring sets of related communication signal streams comprising the steps of analysing the content or parameters associated with a component of one of the signal streams according to a first analysis criteria;
analysing a second component of a related signal stream or parameter associated therewith, according to a second analysis criteria;
providing results of the analysis of the said one of the signal streams and which is responsive to the said analysis according to the second criteria.

This aspect of the present invention therefore advantageously provides for the linking of the analysis of related data streams so as to enhance the analysis of at

least one of the streams.

Advantageously, the said first analysis criteria is arranged to be selected by means of the said second criteria.

Also, the said analysis of the content or parameters associated therewith and the analysis of the signal stream are combined to provide a composite output parameter.

Further, the analysis according to the second criteria occurs prior to the analysis according to the said first criteria.

According to another aspect of the present invention there is provided a communication monitoring system including means for determining an energy envelope representative of at least one communication signal, and means for providing for the subsequent analysis of the said energy envelope.

The monitoring and analysis of the energy envelope represents a particularly efficient and accurate means for determining a variety of aspects and characteristics of, for example, a two-way telephone conversation.

Advantageously, at least two energy envelope files are employed and this can serve to allow for the advantages of stereo recording without disadvantageously doubling storage requirements.

Appropriately, the system can be arranged to allow for the selective analysis of the energy envelope and, in particular, analysis of the energy envelope representative of the final section(s) of, for example, a telephone call/conversation.

Further, the energy envelope may be analyzed so as to identify clipping of the signal which can be indicative of periods of raised voices, or shouting, within the communications traffic stream.

Also, talk/silence ratios can advantageously be determined from the energy envelope so as to identify periods when no communication signals arise, for example, during music-on-hold periods or when a ring-tone is being generated.

The system advantageously further includes storage means for storing the energy envelope for subsequent analysis.

Also, the pattern of activity towards the very end of a call can give indications of abnormal termination - calls being cut-off in the middle of speech or where there is no activity from one or other party for several seconds prior to the end of the call.

A further indication of interest is the speed and clarity of speaking which can be inferred from the gaps between utterances and the average duration of each word spoken.

This aspect of the present invention also advantageously provides for a method of monitoring communication signal including the step of determining an energy envelope representative of at least one communication signal, and subsequently analyzing the said energy envelope. The method can advantageously be conducted in accordance with the system such as that defined above.

According to another aspect of the present invention, there is provided a communications monitoring system including speech recognition means for the identification of words and/or phrases within a communications traffic stream, and

means for varying the scale and/or nature of recognition analysis applied by the speech recognition means.

Advantageously the scale and/or nature of the recognition analysis is arranged to be varied responsive to the identification of at least one party to the communication session. As well as a variety of alternatives, the scale and/or nature of the recognition analysis can be varied on the basis of the length and/or stage of the communication session.

Preferably, the system is arranged to provide speech recognition serving to offer an indication of the level of customer satisfaction. Means can also be provided to generate a score signal indicative of such a level of satisfaction.

Advantageously, separate storage means are provided for storing positive and negative scores.

The system advantageously can also include means for monitoring the operation of a user interface device, the output of which can advantageously be employed in controlling the scale and/or nature of the recognition analysis. For example, since a particular area of interest to a customer can, at any time, be indicated by means of information displayed at a graphical display device associated with the system, a speech recognition module can be operated in a then predetermined manner having regard to the topic being discussed, and thus the keywords and words likely to be spoken.

This further aspect of the present invention also advantageously provides for a method of monitoring a communications traffic stream including the application of

speech recognition to an audio signal arising from part of that communication and having a varying scale and/or nature of recognition analysis and, advantageously, in accordance with the system as defined above.

According to a further aspect of the present invention, there is provided a communications monitoring system including a user interface device for allowing manual input of data to the system, and uses interactions with the said user interface device.

The manner and nature of use of any such user interface device can advantageously provide further information which can usefully be employed in assisting with the monitoring and analysis of the communications traffic stream.

The system can also advantageously allow for monitoring of the accuracy with which a user employs such a user interface device by, for example, monitoring the use of the backspace or delete key of a keyboard etc.

Also, the use of predetermined features of the user interface device can advantageously serve to delineate different sections of the record of use of the user interface device so as to advantageously associate such different sections with respective different sections of the communications traffic stream.

Also, the joint monitoring of the use of the user interface device and the level and/or nature of communications traffic arising can advantageously serve to identify any potential short comings in the skills/efficiencies of the user, for example, from the analysis of, or relation between any pulses arising in the audio signal and corresponding activity noted at the user interface device.

This further aspect of the present invention also provides for a method of monitoring communications signals including the step of monitoring the use of a user interface device associated with the monitoring system.

Advantageously, the invention can also provide for a combination of any one or more of the above-mentioned aspects.

The invention can prove advantageous in at least partially automating the assessment and categorization of recordings. That is, by recording and subsequently analyzing various aspects of the interactions, the system automates the measurement of a range of attributes which previously could only be determined by listening/viewing recordings. For example, these can include, customer satisfaction , call structure (ratio of talking to listening), degree of interruption occurring, degree of feedback given to/from customer, CSR's typing speed and accuracy, CSR's familiarity with and use of the computer application(s) provided, training needs, use of abusive language, occurrence of shouting/heated exchanges, degree of confusion, adherence to script, avoidance of banned words/phrases and the likelihood of customer/CSR having been hung up on.

As a further advantage, the invention can highlight unusual calls for efficient manual review. By measuring the attributes described above, the calls with the highest/lowest scores on each or a combination of such categories can be presented for review. In addition to having automatically selected the calls most likely to be of interest, the present invention provides for mechanisms to present the candidate calls for efficient review. It does this by retaining information, specifically the start and end times related to incidents within the call that led to

it being selected. By way of such information, the most appropriate parts of the call can be selectively played without human intervention. For example, when reviewing potentially abusive calls, only the audio signals approximate to the point where a swear word was identified need be played in order for the reviewer to determine whether or not this is a genuine case of an abusive call or whether a false indication has occurred. Similarly when reviewing calls that are identified as having terminated by one party hanging-up only the last few seconds of the call need be played i.e. the section just prior to the termination. This can therefore allow rapid unattended replay of successive examples without the user having to interact with the system except perhaps to interrupt operation when a potentially interesting call is heard.

The invention can also advantageously offer an objective analysis of the calls. By applying fixed rules and algorithms to the identification of incidents within the calls, and the subsequent categorization or scoring of calls against predetermined criteria and weighting, the scores derived for a given call are deterministic and consistent. Whilst in some respects, the automated scores may not seem as accurate as could be achieved by a well trained human scorer, the fact that the scores can be determined from a much larger sample of, and ideally all available calls, makes them much less subject to random fluctuations than would occur with the small samples such as are scored manually.

Also, the invention can advantageously achieve consistency of analysis. Some aspects of calls that can be measured are independent of the particular products, services or organizations that a customer is dealing with in the interaction. For example, customer satisfaction, if measured

by analysis of the words and phrases spoken by the customer during the call can legitimately be compared across a wide range of organizations and industries. As long as the algorithm used to determine the customer satisfaction rating is kept constant, relative levels of satisfaction can thereby be measured across peer groups and across different industries.

The invention is described further hereinafter, by way of example only, with reference to the accompanying drawings in which:

Fig. 1 is a schematic block diagram of an analysis system embodying the present invention;

Fig. 2 is a schematic block diagram of the recording and analysis sub-system of Fig. 1;

Fig. 3 is a schematic diagram illustrating the separating of an embodiment of the present invention;

Fig. 4 is a schematic diagram illustrating a generic analysis module according to an embodiment of the present invention;

Fig. 5 is a schematic representation of one particular embodiment of analysis module of the present invention;

Fig. 6A and 6B illustrate graphical displays desirable from an analysis module such as that of Fig. 5;

Fig. 7 is a schematic representation of a further embodiment of analysis module embodying the present invention;

Fig. 8 is a schematic representation of yet another

embodiment of analysis module of the present invention;

Fig. 9 is a schematic representation of a further embodiment of analysis module of the present invention; and

Fig. 10 is a schematic representation of yet a further embodiment of analysis module of the present invention.

Turning first to Fig. 1, there is illustrated a multimedia recording and analysis system 14 which is arranged to incorporate the specific methods and systems of the present invention and which is used to monitor the interaction between a CSR and the people/customers and/or systems with whom/which the CSR interacts. Such interaction is conducted by means of a telephone system typically including, as in this example, a console 5 on the CSR's desk and a central switch 8 through which connectivity to the public switched telephone network (PSTN) is achieved via one or more voice circuits 10.

The CSR will typically utilize one or more computer applications accessed through a terminal or PC at their desk 2 with which they can interact by means of the screen 1, mouse 3 and a keyboard 4. The software applications employed may run locally on such a PC or centrally on one or more application server(s) 7.

The system 14 embodying aspects of the present invention can advantageously offer connections to so as to monitor and/or record any required combination of aspects such as the audio content of telephone conversations undertaken by the CSR's by means of a speech tap 13 or the contents of the screen on the CSR's desk during interactions. This latter aspect can require software to be loaded onto the CSR's PC 2 to obtain such information and pass it to the recording system.

The keystrokes, mouse movements and actions of the CSR can be monitored and can also typically require software to be loaded onto the CSR's PC 2. Monitoring of the context of, and data entered into or displayed by, the applications being used can typically require system integration to have the applications pass such information to the recording system. Further, the details of calls placed, queued and transferred etc by the telephony system can be monitored.

It should of course be appreciated that this is merely a typical example of a system that can employ the present invention and numerous variants on this theme are well known, such as the use of Voice Over IP (VoIP), the tapping of the audio signals at the console rather than the trunks, and the use of silent monitoring features to allow for tapping into selected consoles.

Fig. 2 represents a high level view of the major components within a recording and analysis system embodying the present invention. It should be appreciated that such systems are typically deployed across multiple sites and are implemented on multiple computer platforms. However, the major functional blocks remain the same.

Incoming data streams, such as voice, screen content, events etc. 17, and recording control information such as CTI information from an ACD that trigger commands from a desktop application etc. 18 are processed by one or more record processors 19. The net results of such processing are, first the storage of call content in some form of non-volatile storage system such as a disk file storage system 16 can be achieved, and secondly details about the recordings made can be stored in a relational database 15 allowing subsequent search and retrieval on the basis of a number of criteria.

These recordings and the details about them are then subsequently available for one or more search and replay applications that allow users, or other applications such as a customer relationship management system, to access the particular recording(s) they require. One such application is a quality assessment application which will typically make random or pre-programmed selections of recordings and present these to a reviewer for evaluation and subsequent analysis of the results of the said evaluations. Such call-flow recording is described in the present applicant's International application WO 01/52510.

The enhancements to such recording and analysis systems that can be achieved by the present invention relate to methods that act upon the content of the recordings and/or the details about such recordings. In a system of the type shown in Fig. 2, such processes may be advantageously applied at one or more of a variety of points in the system. The optimum location for each method will depend generally on the analysis being performed and also the accuracy required and the topology of the system.

Examples of the options for deploying such methods as described are shown in Fig. 3 and are as follows. First, the point 20 in the system at which the method is employed can comprise part of the record process, with access to the raw, unprocessed, information as received. This may prove to be the only way in which to influence the operation of the recording system as a result of the analyses performed in real-time. Such a location may also be the only point at which unadulterated information is available, for example un-compressed audio that is only stored subsequently to disk once it has been compressed and/or mixed with information/data from other input channels. An alternative location comprises the point at which data is written to

disk. This can prove particularly useful if only a subset of input data is to be recorded. By applying the required algorithms to data at this point, resources are not wasted in attempting to process data that was not actually recorded.

The present applicant's European patent application EP-A-0 833 489 discloses features such as those described above.

A further option comprises a location 22 forming part of an offline process. Here, although the overhead of having to query the database and/or read the recording content from disk is incurred, this does allow ongoing 24 hour analysis since it may not prove possible to keep up with the rate of recordings made during the busiest periods of the day.

Advantageously, it can be arranged that such analysis modules are deployed on the CSR'S desktop PCS during periods when the PCS would otherwise be idle. This allows economic deployment of complex analysis such as full speech recognition which, otherwise, would disadvantageously require additional investment in additional processors or would have to be restricted to a much smaller subset of the total recordings.

Also, at location 23, some of the analyses may be performed as part of search and replay applications. This is particularly advantageous for analyses that can be performed rapidly on a small set of calls that are already known to be of interest to the application/user in question. The details about the recordings, and the recordings themselves, will in some instances, already have been retrieved by that application and so will be accessible to the analysis tools of the present invention.

Such an arrangement is illustrated by reference to Fig. 4 in which call recording details 25 and call recording content 26 are input to an analysis module 24.

The analysis algorithm within the module 24 operates on these inputs to produce further details 27 about the recordings and/or further recordings 28 derived from the input recordings.

Again, according to deployment, these outputs may simply be used by the application holding the module such as at location 23 or may be written back to the database 15 and/or file system 16. In the latter case it should be noted that the outputs from any module are therefore available as inputs to other modules allowing cascading of analyses such that some modules may produce interim results whilst others further process the outputs or combine them with the outputs of still further modules to produce composite and derived outputs.

An example of such a module is shown in Fig. 5 and is arranged to produce an output file for each input voice recording which summaries the audio level or energy present throughout the recording. In its simplest form such an energy envelope module 33 is arranged to operate on an incoming audio signal 30 and convert it to a signed linear encoding scheme 31 if it is not already in such a format. It then averages the absolute value (or, optionally the square of the value) over a fixed interval in the order of typically 50ms. This interval is chosen so that when displayed graphically, the resolution of the samples is sufficient to allow easy visibility of the words and pauses in the recording. An example of a graphical output derived from such an energy envelope' file is shown in Figure 6A

These files prove useful in serving as thumbnail graphical overviews of calls as well as serving as useful input for subsequent analysis stages. The energy files avoid the need to retrieve the entire audio recording, and to decompress it, and so make many subsequent analyses viable that would otherwise prove prohibitive due to network bandwidth and/or processing requirements.

As with all other parameters recorded in the invention, the storage may be beneficially accomplished by writing the information in the form of an XML file. The structure of the energy envelope file can be very simple for example it can comprise merely a succession of the average energy values. Beneficially however, the maximum energy value encountered within the file is noted at the start of the file. This allows an application using this file to perform scaling on the file without first having to read the entire file in search of the maximum value.

This maximum value is noted by a statistical analysis function 36 as illustrated in Fig. 5 as the recording is being processed. Additional statistics derived from the energy values may also be derived at this time. For example, the ratio of quiet periods (when energy is below a specified threshold for a high proportion of the samples) to active periods can be obtained. Also, the prevalence and location within the call of any periods of clipping, i.e. where the audio signal saturates at the extreme of the available audio range leading to distortion can be identified. This may indicate extreme volume levels such as those arising due to the customer shouting.

This module is advantageously deployed where the audio signal has not yet been compressed. It is much more economical to convert standard telephony signals (e.g. in

G.711 mu-law or A-law) to linear than it is to decompress a heavily compressed audio signal.

Furthermore, some parameters such as the "degree of clipping" are adversely affected by the compression algorithms employed.

The module 33 can further advantageously be deployed prior to any mixing of audio signals such as occurs at location 20 in Fig. 3 such that the output energy envelope file reflects the audio levels in a single direction of speech received. Thus, although the original transmit and receive signals may be subsequently mixed into a single audio file for more efficient storage, the two energy envelope files may be used to produce a clear graphical display as shown in Figure 6B that highlights who was talking and when, and also enables interruptions to be highlighted as indicated by arrows A.

Referring now to Fig. 7, an energy envelope analysis module 37 takes as its input, one or more energy envelope files 38 plus details about the calls 40 to which they relate. Typically, the module 37 will serve to analyze the two energy envelope files relating to the transmit and receive audio paths for a single call but may also compare a set of energy envelope files for a set of supposedly similar calls. Statistical analysis indicated at 39, 41 of the input energy envelope files can be performed to derive output information such as those discussed as follows.

The proportion of talk periods to listen periods within the call. The frequency of confirmatory feedback from each party in the call, i.e. when one party is speaking, the other will normally respond with an 'uh-huh' or similar utterances which shows as a brief burst of energy on one channel in the midst of a sustained burst of energy due to the sentence

being spoken on the other. The frequency and proportion of argumentative interruption which can be defined as sustained activity on both channels concurrently for a period exceeding the normal time needed for one party to concede control of the conversation to the other. The proportion of silent periods within the call. The locations of sustained silences within the call and also which party eventually breaks the silence. An unusual call termination pattern different from the usual pattern at the end of a call, when each party speaks briefly, to say goodbye etc. followed by a brief pause and then a loud click as the call is terminated. An abrupt termination of a call within a sustained period of activity by one or other party which can indicate a likely abnormal call-termination. Episodes of shouting or increasing volume, in which the average volume of one or both speakers alters during the course of the call and which can be flagged as a possible indication of a heated conversation.

Any of the aforementioned may be combined with a weighting profile that influences the effect of each function of time throughout the call. For example, the determined value of output information preferred talk-to-listen profile may be 50:50 during the first 30% of the call but then may change to a ratio 30:70 thereafter.

A more sophisticated analysis can be performed by utilizing speech recognition tools in order to identify keywords within recordings or to perform large vocabulary transcription of the audio into text. Fig. 8 illustrates such a module 43. The audio streams and, optionally, energy envelope files previously generated 44 are used as an input, along with any pre-existing details 45 about the recordings. The input may initially be sliced at location 46 using the energy envelope files and other details to determine which

portions, if not all, of the recording are to be analyzed by a speech recognition engine 47. The recognition engine 47 is then delivered to a database 49 of entries listing transcript and/or individual words recognized via a database and/or to a file 50 holding similar details directly on the file storage system.

The output from such a speech recognition module 43 is typically one or more of a so-called best guess transcription of the call, or a sequence of recognized words or phrases, their locations within the call and some measure of the likely degree of confidence in their recognition

Such details can be stored for direct searching so as, for example, to find all calls containing a specific word or for further analysis.

The speech recognition module 43 is advantageously deployed where the audio signal has not yet been compressed. Recognition accuracy and the ease of computation are found to be better for an un-compressed signal than for a compressed one.

The speech recognition module 43 is further advantageously deployed prior to any mixing of the audio signals such as at location 20 so that a single speaker can be recognized at a time. This allows the optional deployment of speaker specific recognition models where the speaker is known from the recording details and also ensures that the output is unambiguously linked to the appropriate party to the call. Hence the output is both more accurate and more useful.

Advantageously, if the unmixed stereo recording is unavailable, the speech recognition module 43 may take as its inputs, the mixed audio recording and the energy

envelope files previously generated. These advantageously allow the recognition engine to :-

a) determine which of the two speakers on the call is active at any time and hence apply the most appropriate speaker and vocabulary model enhancing accuracy;

b) label the output with a clear indication as to which party uttered the words detected; and

c) identify more clearly the start and end of utterances which otherwise may merge into one and hence result in lower recognition accuracy as the recognition engine expects a single phrase or sentence in each contiguous utterance rather than two sentences.

Advantageously, the recognition engine 47 is instructed to recognize less than the entire call. As recognition is extremely processor intensive, it can prove beneficial to analyze selected portions of the call. For example, the first 30 seconds can be analyzed to determine the type of call, and the last 30 seconds analyzed to determine the outcome of the call and level of customer satisfaction.

Further, the above partial analysis of the call may be optimized by using the previously derived energy envelope files. Using these energy envelope files, the location and duration of the first and last n utterances by each party can easily be determined and the recognition engine directed to process only these portions of the call. For example, by analyzing the last utterance made by each party it is normally possible to determine the appropriateness of the closure of the call and hence to identify those in which an unusual call closure occurred such as when a CSR hung-up on a customer.

Advantageously, the speech recognition module 43 may only be instructed to analyze a subset of calls that have already

proved to be of potential interest due to any combination of recording details and details derived for example from prior energy envelope analysis.

The speech recognition module 43 may also make use of a call flow recording input that indicates the current context of the application(s) being used by the CSR. A vocabulary and grammar model applied by the recognition engine can be influenced by a determination of which application/form and field is active on the CSR's screen. This leads to more accurate recognition and the context can be recorded along with the transcript output allowing subsequent modules to search for words uttered at specific points in the structure interaction flow.

Turning now to Fig. 9, there is illustrated a language analysis module 51 which is used to process the output of the speech recognition module. By comparing the words identified in calls 53 against a list of phrases that are of interest, the call can be annotated with database entries 57 and/or additional recorded file information 58 that highlight the presence or absence of these phrases. The output typically includes the start position of the phrase, its duration and confidence of recognition allowing subsequent review of exactly this portion of the call.

Advantageously, the phrases being sought may include wildcard words, e.g. for example the phrase "you've been? helpful" would match a phrase that contained any word between "been" and "helpful".

The phrases can be grouped according to the type of information they indicate. For example, the above phrase would be a customer satisfaction indicator phrase, whereas the "I'm not sure" would be identified as a training need

indicator etc.

Further, each phrase can be allotted a score as to how relevant it is to the type of information sought. For example a simple "thank you" could score +0.1 on the customer satisfaction indicator category whereas "thank you very much" would score +0.3. By storing these relative scores the reviewer can see the relative importance of each phrase matched when reviewing calls.

Advantageously, any cumulative score achieved by each call on each of the categories is summed by a score accumulator 55 and the net results for each call are stored to the database 57 and/or the file 58. The score accumulator may apply a time function that weights the scores for specific categories according to the time within the call, whether absolute or relative, that the phrase is recognized. For example, the customer satisfaction indicator would be weighted more heavily towards the end of the call rather than the beginning as the customer may already be happy or upset due to other factors at the start of the call. The success of the call is more accurately determined by the customer's state at the end of the call.

In situations where both positive and negative scores are assigned to phrases in the same category the system is arranged to separate total positive and negative scores, rather than merely seeking to cancel these out. A call with extremes of positive and negative satisfaction is naturally of more interest and different from one where no expression of satisfaction is made.

Advantageously, the language analysis module 51 may also make use of Call Flow Recording input that indicates the current context of the application(s) being used by the CSR.

Phrases and their scores can be linked to specific contexts within the application and their scores and applicability varied according to this context.

The module 51 may further, also serve to operate on the output of a keystroke analysis module such as described below, taking the words entered into the computer system as another source of input on which phrase matching and scoring can be performed.

With reference now to Fig. 10, there is illustrated a keystroke/mouse analysis module 59 that analyses the screen content and/or keystroke/mouse recordings that can be made at a CSR's PC. Three independent analyses of the keystrokes provide for the following.

First, word and phrase identification 62 can be achieved by combining successive keystrokes into words and then phrases since the module can make the keystroke information a useful search field. The module 59 must take account of the use of mouse clicks, tab keys, enter key etc. that delimit the inputs into a specific field and hence separate subsequent text from that entered prior to the delimiter. Interval analysis 64 is achieved by analyzing the time between successive keystrokes.

Secondly, an indication of typing skills can be obtained. The use of specific keys such as backspace and delete can also give indications of level of typing accuracy. The results of this analysis are useful in targeting typing training courses at those most likely to benefit.

Finally, a range analysis function 65 can be achieved by noting the variety of keys used and compared against other calls. It is then possible to identify users who are

unfamiliar with, for example, standard windows shortcut keys (Alt+C) or application specific shortcuts (F2 for order form). The frequency of use of these less common keystrokes can be stored and subsequently used to identify opportunities for windows and/or application specific training.

The outputs of the above stages may be accumulated at location 66 through the call and the net results stored in addition to the individual instances.

The output of this module can again comprise database entries 68 for the call and/or file content 69 listing the results of the analyses 62,64,65,66 discussed above.

CLAIMS

1. A method of monitoring sets of related communication signal streams comprising the steps of analysing the content or parameters associated with a component of one of the signal streams according to a first analysis criteria;
analysing a second component of a related signal stream or parameter associated therewith, according to a second analysis criteria;
providing results of the analysis of the said one of the signal streams and which is responsive to the said analysis according to the second criteria.

2. A method as claimed in Claim 1, wherein the said first analysis criteria is selected by means of the said second criteria.

3. A method as claimed in Claim 1, wherein the said first analysis criteria is arranged to be adapted by means of the said second criteria.

4. A method as claimed in Claim 1, 2 or 3, and wherein the said analysis of the said content or parameters and the analysis of the signal stream are combined to provide a composite output parameter.

5. A method as claimed in any one or more of Claims 1-4, wherein the analysis according to the second criteria occurs prior to the analysis according to the said first criteria.

6. A method as claimed in any one or more of Claims 1-5, and including the step of recording the signal stream.

7. A method as claimed in any one or more of Claims 1-6, and including the step of introducing timing information serving to locate analysed portions within the signal stream.

8. A communications monitoring system and including means for executing the method of any one or more of Claims 1-7.

9. A communication monitoring method including the steps of determining an energy envelope representative of at least one communication signal, and providing for the subsequent analysis of the said energy envelope.

10. A method as claimed in Claim 9, wherein at least two energy envelope files are employed.

11. A method as claimed in Claim 9 or 10, and arranged to allow for the selective analysis of the energy envelope.

12. A method as claimed in Claim 11, and arranged to allow for analysis of the energy envelope representative of the final section of the communication signal.

13. A method as claimed in Claim 9, 10, 11 or 12, and including the step of analysing the energy envelope so as to identify clipping of the signal.

14. A method as claimed in Claim 9, 10, 11, 12 or 13, and including the step of determining sound/silence ratios from the energy envelope.

15. A method as claimed in any one or more of Claims 9 to 14, and including the step of analysing the duration of sound passages.

16. A method as claimed in any one or more of Claims 9 to 15 and including the step of analysing the delays between signal transmissions in different directions.

17. A method as claimed in any one or more of Claims 9 to 16, and including the step of storing the energy envelope for analysis.

18. A method of monitoring a communication signal as defined in any one or more of Claims 1 to 7 and including the method steps of any one or more of Claims 9 to 17.

19. A communications monitoring system and including means for executing the method as defined in any one or more of Claims 9 to 17.

20. A communications monitoring method including the steps of conducting speech recognition for the identification of words and/or phrases within a communications traffic stream, and including the step of varying the scale and/or nature of recognition analysis applied for the speech recognition responsive to the analyses of content or parameters associated with the communications stream or related streams.

21. A method as claimed in Claim 20, wherein the scale and/or nature of the recognition analysis is arranged to be varied responsive to the identification of at least one party to the communication session.

22. A method as claimed in Claim 20, wherein the scale and/or nature of the recognition analysis is arranged to be varied on the basis of the length and/or stage of the communication session.

23. A method as claimed in Claim 20, 21 or 22, and including the step of generating a score signal indicative of such a level of satisfaction

24. A method as claimed in Claim 20, 21, 22 or 23, and including the step of monitoring the operation of a user interface device, the output of which is employed in controlling or adapting the recognition analysis.

25. A communication monitoring method of any one or more of Claims 1 to 7 and 9 to 18, and including the steps of Claims 20 to 24.

26. A communications monitoring system and including means for executing the method steps of any one or more of Claims 20 to 25.

27. A communications monitoring method including the step of monitoring usage of a user-interface device associated and arranged to be used concurrently, with the communication stream and controlling the communications monitoring responsive to the results of said monitored usage.

28. A method as claimed in Claim 27, and including the step of monitoring the accuracy with which a user employs the said interface device.

29. A method as claimed in Claim 27 or 28 and wherein the said user-interface device comprises a computer device.

30. A method as claimed in Claim 29 and including the step of monitoring the keystrokes and/or mouse actions of the user.

31. A method as claimed in Claim 29 or 30 and including the steps of monitoring the applications, documents and/or windows selected by the user.

32. A method as claimed in any one or more of Claims 27 to 31, and including the step of delineating different sections of a record of use of the said interface device so as to associate such different sections with respective different sections of the monitored communication.

33. A method as claimed in any one or more of Claims 27 to 32, and including the step of monitoring jointly the use of the said interface device and the level and/or nature of communications traffic to identify characteristics of the user.

34. A communications monitoring method of any one or more of Claims 1 to 7, 9 to 18, 20 to 24 and including the steps of any one or more of Claims 26 to 33.

35. A communications monitoring system including the means for executing the method steps of any or more of Claims 27 to 34.

FIG. 1

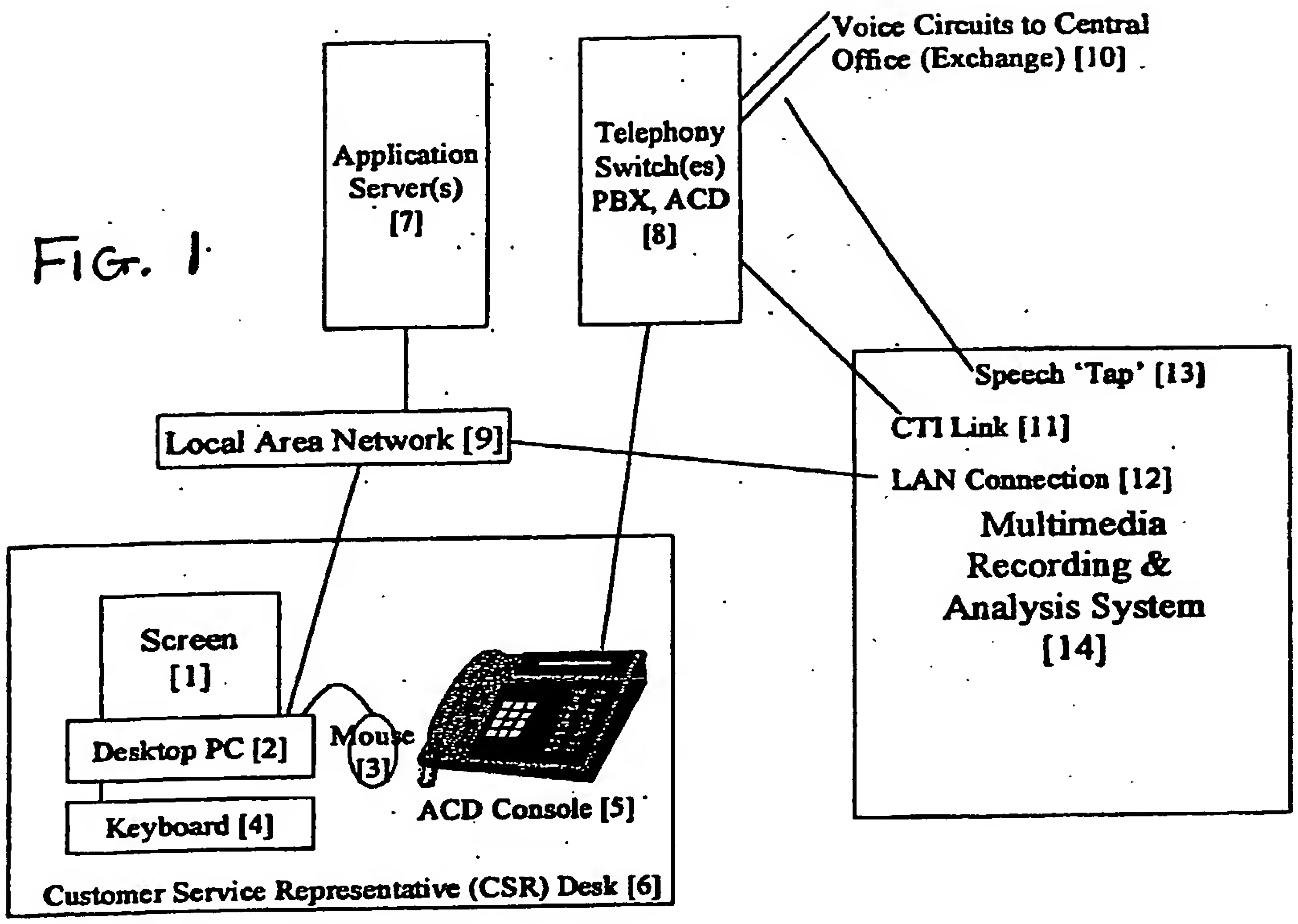
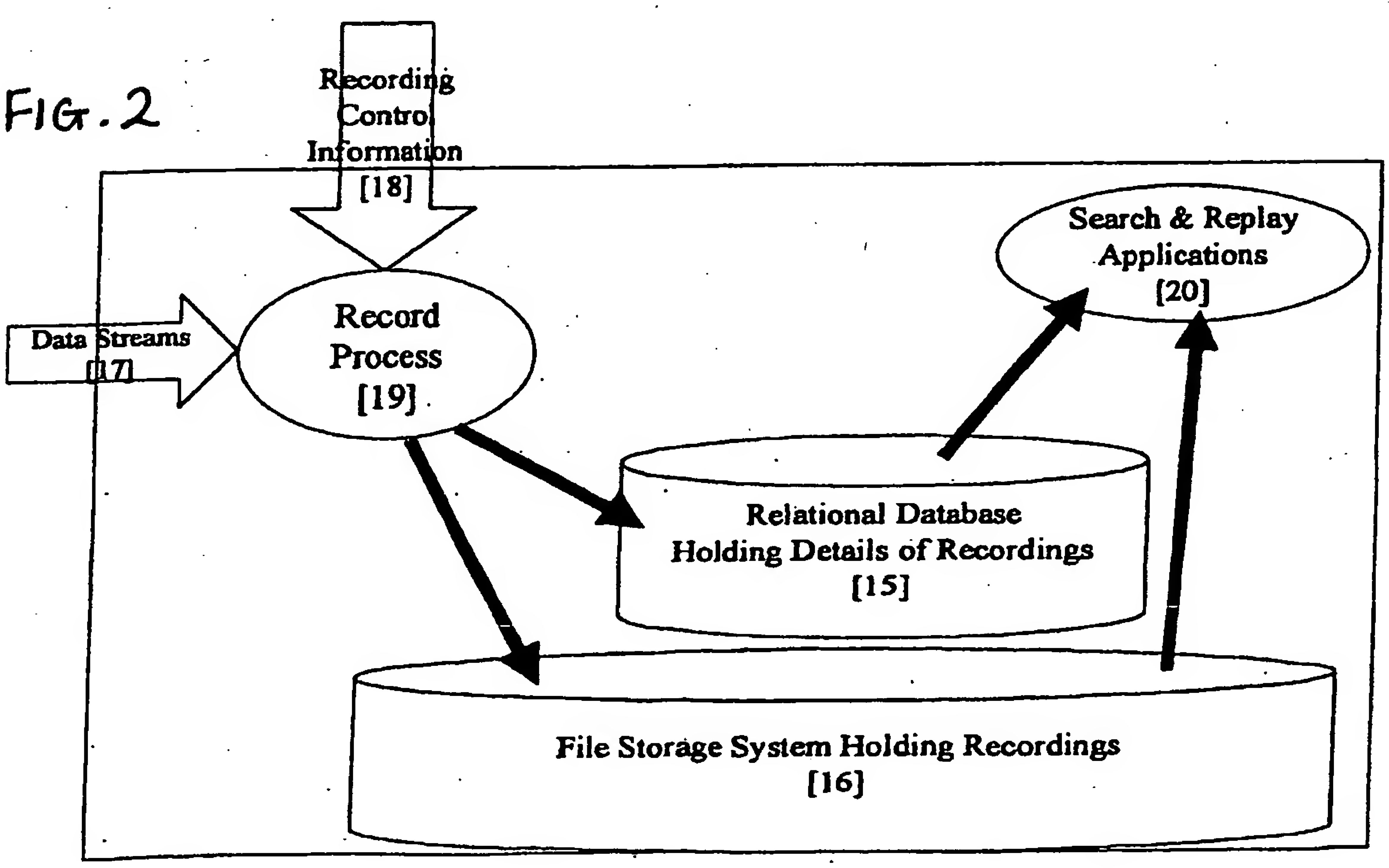

Application Server(s) [7]
Telephony Switch(es) PBX, ACD [8]
Voice Circuits to Central Office (Exchange) [10]
Speech 'Tap' [13]
CTI Link [11]
LAN Connection [12]
Multimedia Recording & Analysis System [14]
Local Area Network [9]
Screen [1]
Desktop PC [2]
Mouse [3]
Keyboard [4]
ACD Console [5]
Customer Service Representative (CSR) Desk [6]

FIG. 2


Recording Control Information [18]
Data Streams [17]
Record Process [19]
Search & Replay Applications [20]
Relational Database Holding Details of Recordings [15]
File Storage System Holding Recordings [16]

BEST AVAILABLE COPY

FIG. 3

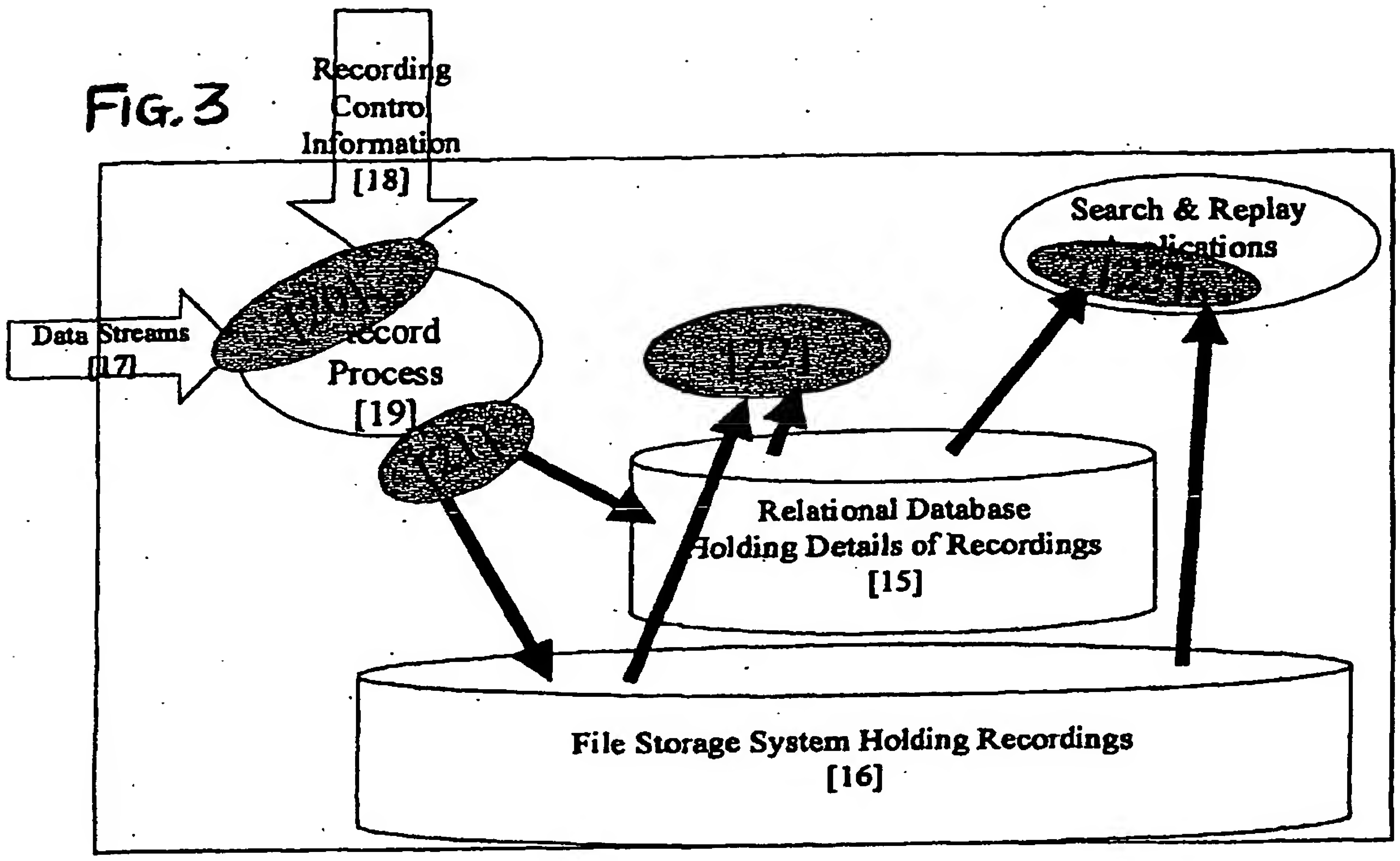
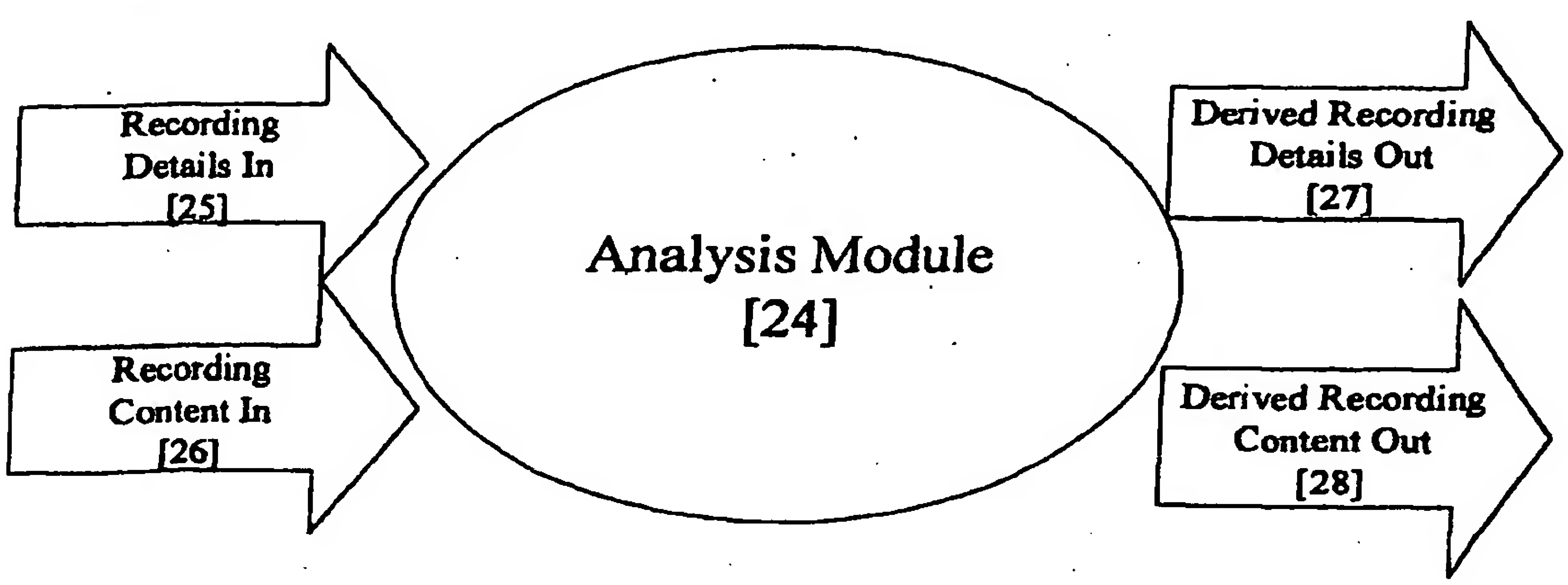


FIG. 4



BEST AVAILABLE COPY

FIG. 5

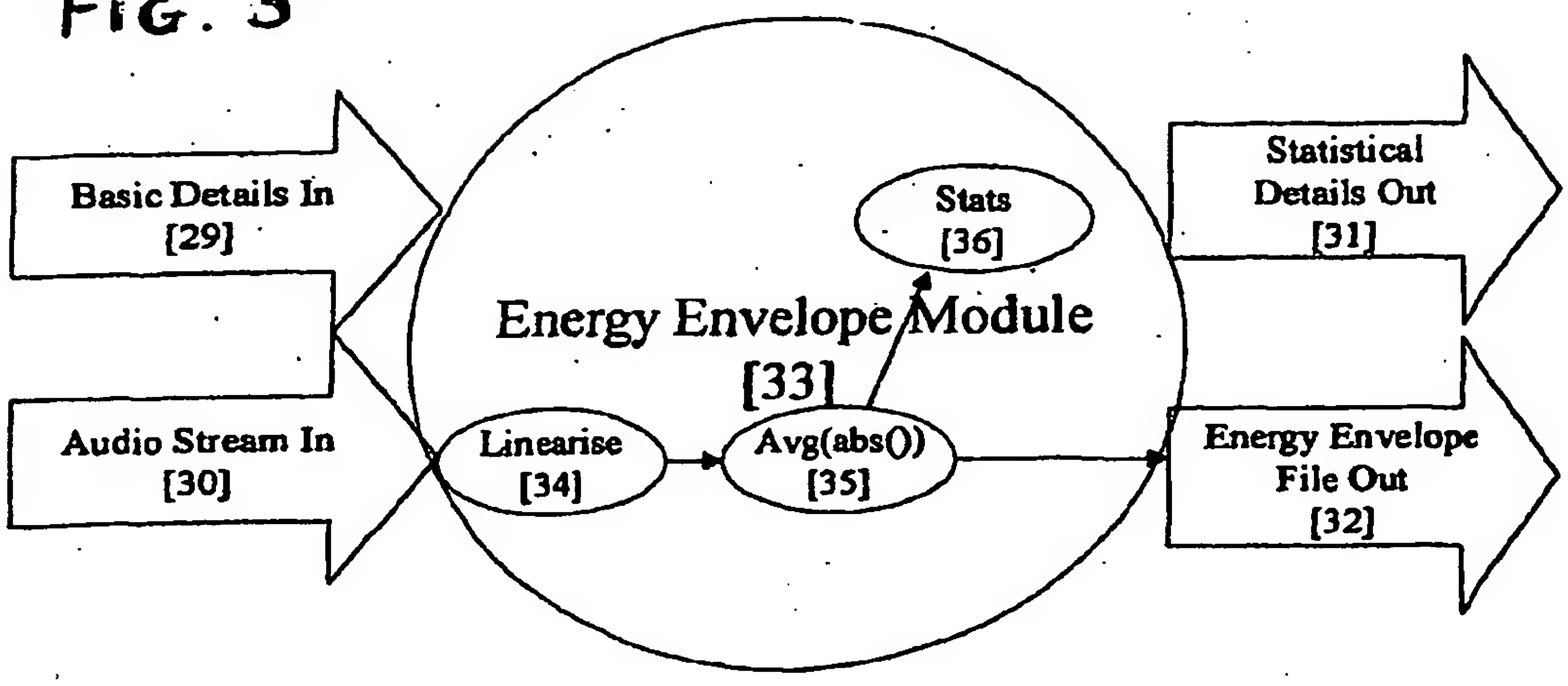


FIG. 6A

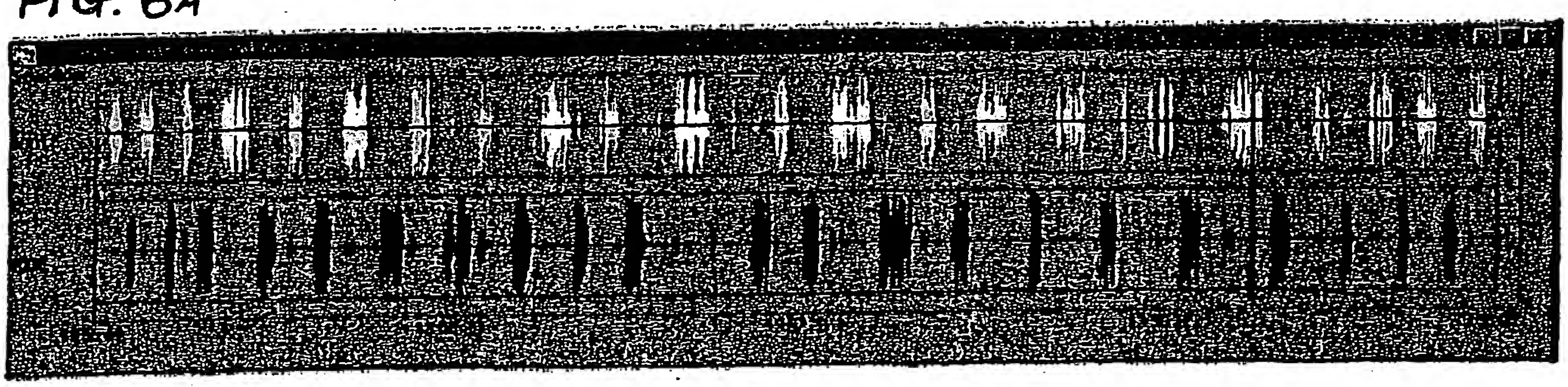


FIG. 6B

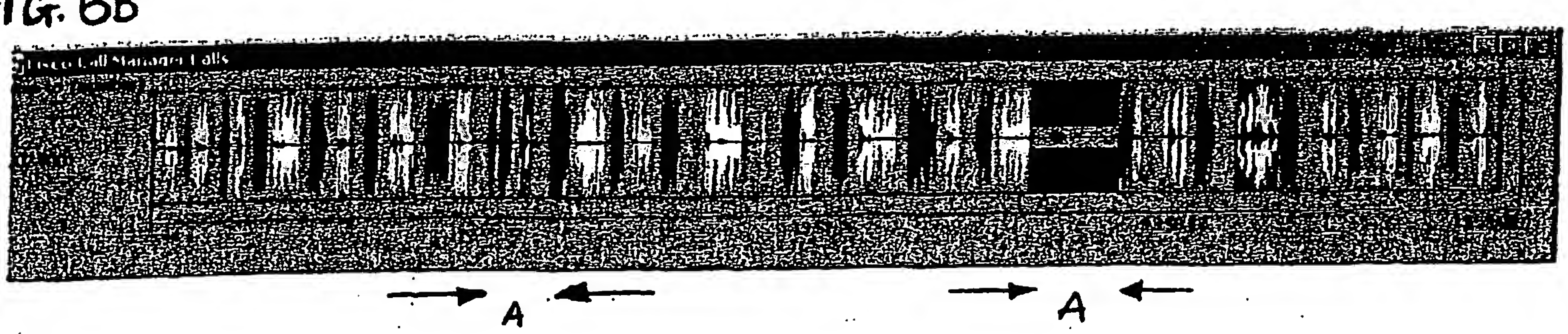


FIG. 7

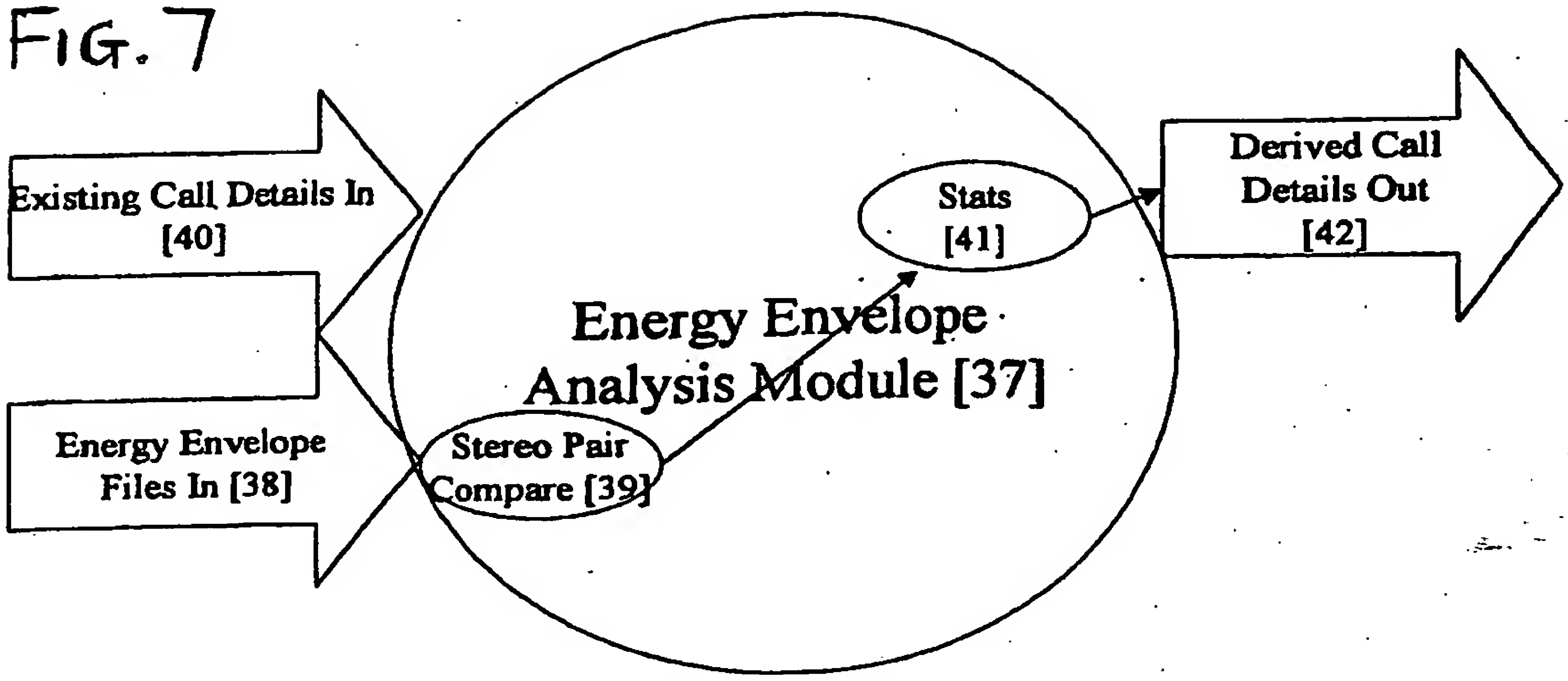

Existing Call Details In [40]
Energy Envelope Files In [38]
Stats [41]
Energy Envelope Analysis Module [37]
Stereo Pair Compare [39]
Derived Call Details Out [42]

FIG. 8

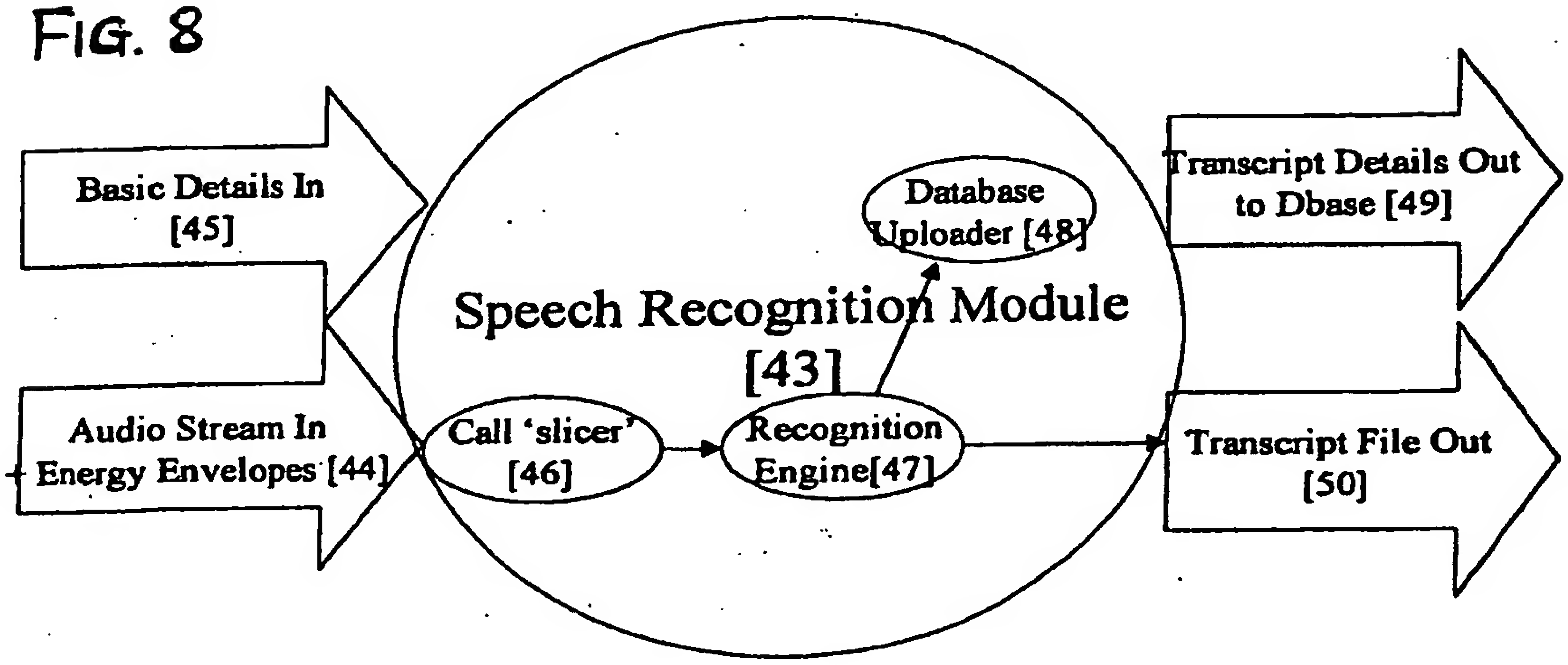

Basic Details In [45]
Audio Stream In Energy Envelopes [44]
Database Uploader [48]
Speech Recognition Module [43]
Call 'slicer' [46]
Recognition Engine[47]
Transcript Details Out to Dbase [49]
Transcript File Out [50]

FIG. 9

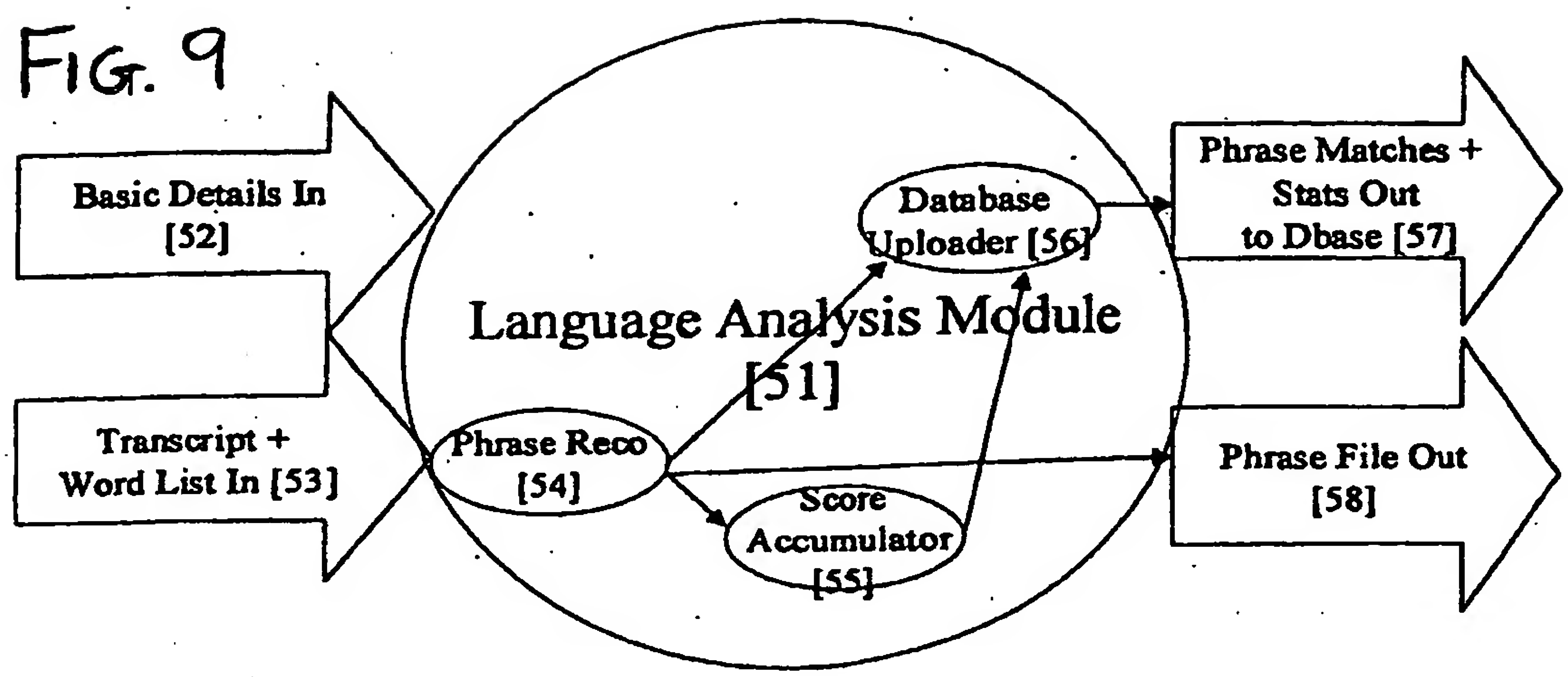
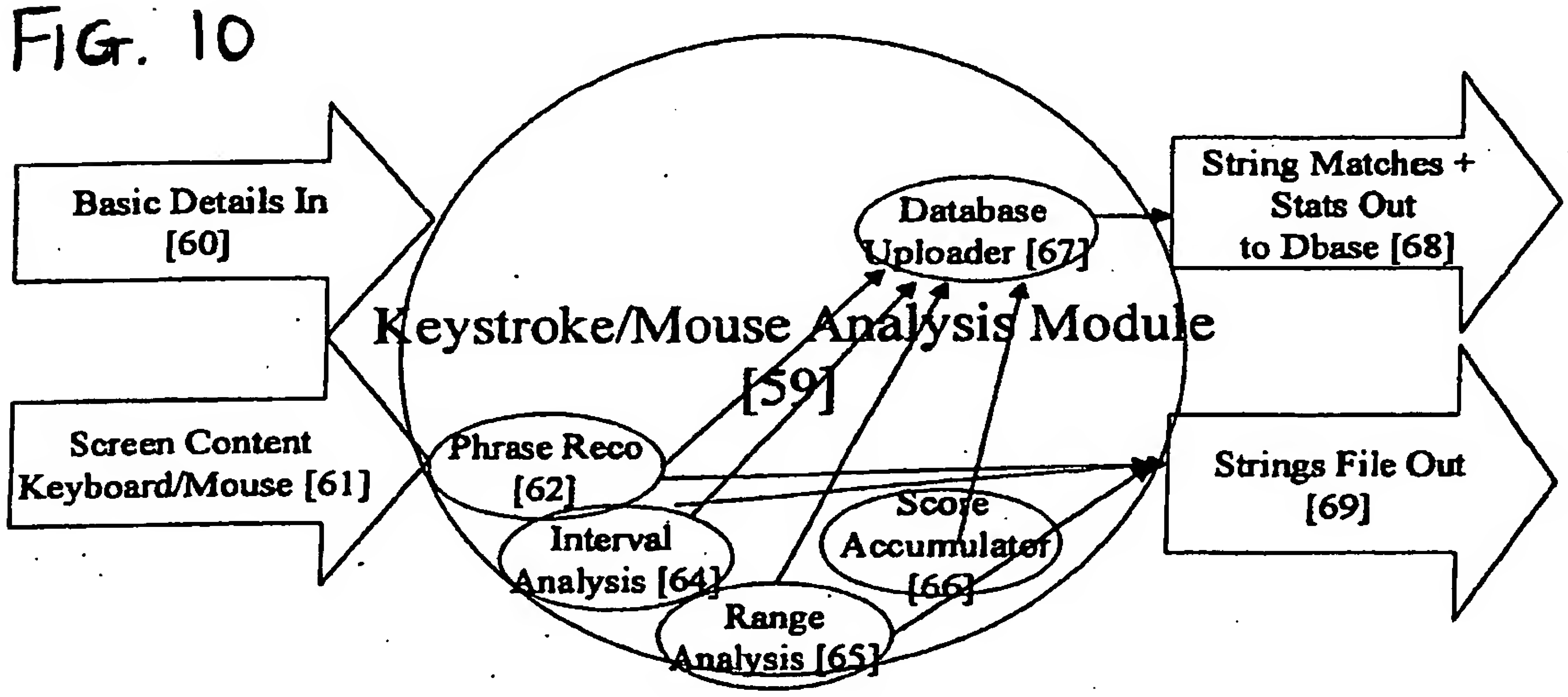

Basic Details In
[52]
Transcript +
Word List In [53]
Language Analysis Module
[51]
Database
Uploader [56]
Phrase Reco
[54]
Score
Accumulator
[55]
Phrase Matches +
Stats Out
to Dbase [57]
Phrase File Out
[58]

FIG. 10


Basic Details In
[60]
Screen Content
Keyboard/Mouse [61]
Keystroke/Mouse Analysis Module
[59]
Database
Uploader [67]
Phrase Reco
[62]
Interval
Analysis [64]
Range
Analysis [65]
Score
Accumulator
[66]
String Matches +
Stats Out
to Dbase [68]
Strings File Out
[69]

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/067360 A2

(51) International Patent Classification[7]: G06F

(21) International Application Number: PCT/IL02/01042

(22) International Filing Date: 26 December 2002 (26.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/354,209 6 February 2002 (06.02.2002) US

(71) Applicant (*for all designated States except US*): **NICE SYSTEMS LTD.** [IL/IL]; Hapnina Street 8, 43107 Ra'anana (IL).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **ELAZAR, Avishai** [IL/IL]; Brandes Street 19/8, 43357 Ra'anana (IL). **BACK, Ayelet** [IL/IL]; Haronson Street 12/4, 68012 Tel Aviv (IL). **DVIR, Igal** [IL/IL]; Akiva Street 86, 43263 Ra'anana (IL). **KOREN-BLUMSTEIN, Guy** [IL/IL]; Emek Ayalon Street 14, 44627 Kfar Saba (IL).

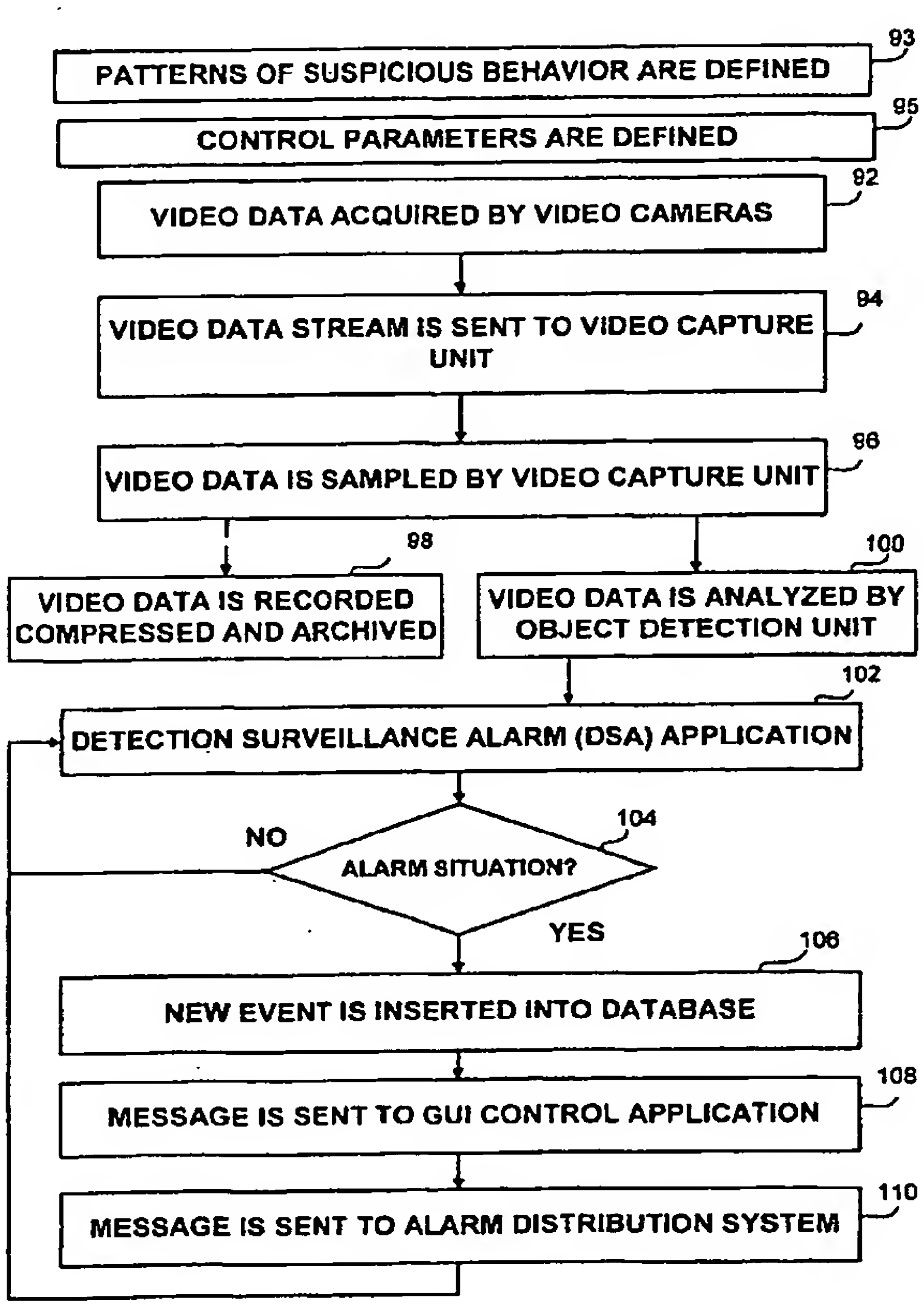
(74) Agents: **AGMON, Jonathan** et al.; Soroker-Agmon, Advocates and Patent Attorneys, Levinstein Tower, 12th Floor, Petach Tikva Road 23, 66184 Tel Aviv (IL).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR VIDEO CONTENT ANALYSIS-BASED DETECTION, SURVEILLANCE AND ALARM MANAGEMENT



(57) Abstract: A surveillance system and method for the automatic detection of potential alarm situation via a recorded surveillance content analysis and for the management of the detected unattended object situation. The proposed system and method is operative in capturing surveillance content, analyzing the captured content and providing in real time a set of alarm messages to a set of diverse devices. The system provides event based debriefing according to captured objects captured by one or more cameras covering different scenes. The invention is implemented in the context of unattended objects (such as luggage, vehicles or persons), parking or driving in restricted zones, controlling access of persons into restricted zones, preventing loss of objects such as luggage or persons and counting of persons.

WO 03/067360 A2

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SYSTEM AND METHOD FOR VIDEO CONTENT ANALYSIS-BASED DETECTION, SURVEILLANCE, AND ALARM MANAGEMENT

BACKGROUND OF THE INVENTION

RELATED APPLICATIONS

The present invention relates and claims priority from US provisional patent application serial number 60/354,209 titled ALARM SYSTEM BASED ON VIDEO ANALYSIS, filed 6 February 2002.

FIELD OF THE INVENTION

The present invention relates to video surveillance systems in general, and more particularly to a video content analysis-based detection, surveillance and alarm management system.

DISCUSSION OF THE RELATED ART

Due to the increasing number of terror attacks and potential terror-related threats, one of the most critical surveillance challenges today is the timely and accurate detection of suspicious objects, such as unattended luggage, illegally parked vehicles, suspicious persons, and the like, in or near airports, train stations, federal and state government buildings, hotels, schools, crowded public places typically situated at city centers, and other sensitive areas. In accordance with the prevailing known tactics of terrorist organizations, unattended innocent-looking objects, such as a suitcase, could contain hidden explosive materials installed therein to effect a controlled explosion and thereby inflict massive impact damage to the near environment and the individuals within. Likewise recent attacks have been perpetrated through the use of vehicle bombs, seemingly innocent until explosion. Since these hidden explosives are usually activated by the setting of carefully timed (typically short-period) detonator means, or by an operator who is at the scene or close by, the prompt, rapid and timely detection of suspicious objects, such as unattended luggage, vehicles parking in forbidden zones,

suspicious persons, persons leaving unattended suitcases or vehicles, and the like, could prevent life-threatening situations. Similarly, it is important in areas, such as airports, to be able to track persons and objects, such as suitcases and cars, to assist in locating lost luggage, and to restrict access of persons or cars to certain zones. The applications of such abilities are not only for security purposes.

Recently, the authorities responsible for the safety of the public have been attempting to cope with the problems listed above in the most obvious manner by increasing the number of human personnel tasked for the detection, identification and consequent handling of suspicious objects, including vehicles, luggage and persons. At the same time, in order to maintain substantially unobstructed passenger flow and in order to minimize transport delays and consequent public frustration, the security personnel have been obliged to utilize inefficient and time-consuming procedures. One drawback of the above human-centric solution concerns the substantially increased expenses associated with the hiring of a large number of additional personnel. Another drawback concerns the inherent inefficiency of the human-centric procedures involved. For example, specific airport security personnel must perform visual scanning, tracking, and optional handling of objects in sensitive transit areas 24 hours a day, where the sheer number of luggage passing through these areas effect increased fatigue accompanied naturally with diminished concentration. In the same manner, in a traffic-extensive area wherein specific security personnel must watch, track and optionally handle vehicles parking in restricted areas, natural weariness soon sets in and the efficiency of the human-centric procedure gradually deteriorates.

Currently available surveillance systems are designed for assisting human security officers. These systems typically include various image acquisition devices, such as video cameras, for capturing and recording imagery content, and various detector devices, such as movement detectors. The existing surveillance systems have several important disadvantages. The type of alarms provided by the detectors is substantially limited. The video images recorded by the cameras are required to be monitored constantly by human security personnel

in order to detect suspicious objects, people and behavior. An alarm situation has to be identified and suitably handled by the personnel where a typical handling activity is the manual generation and distribution of a suitable alarm signal. Since these surveillance systems are based on human intervention the problems related to natural human-specific processes, such as fatigue, lack of concentration, and the like, are still remain in effect.

A further drawback of existing surveillance systems concern the failure of those systems to handle certain inherently suspicious events that were captured by the cameras monitoring a scene. For example, current surveillance systems associated with airport security application, typically fail to identify a situation as suspicious where the situation involves a vehicle arriving at a monitored airport terminal, an occupant of the vehicle leaving of the vehicle, and the departure of the occupant from the monitored scene in a direction that is opposite to the terminal.

Yet another drawback of the current systems concern the inability of the current systems to identify a set of events linked to the same object in the same area throughout a pre-defined surveillance period. For example, when "suitcase" object is left in the scene by a first person and later it is picked up by a second person then the leaving of the suitcase and the picking up of the suitcase constitute a set of linked events.

Still another drawback of the current systems concern the inherent passivity of systems due to the fact that the operations of the systems are based on events initiated by the operators and due to the fact that the systems provide no built-in alerts.

In addition, existing systems are incapable of associating a retrieved event or object through the use of important parameters, such as color of hair, color of clothing and shoes, complexion (via the use of a color histogram), facial features (via face recognition routines), normalized size of the object (distance from the camera), and the like.

It would be easily perceived by one with ordinary skills in the art that there is a need for an advanced and enhanced surveillance, object tracking and identification system. Such a system would preferably automate the procedure concerning the identification of an unattended object substantially and would utilize cost-effective, efficient methods.

SUMMARY OF THE PRESENT INVENTION

One aspect of the present invention relates to a method for analyzing video data, comprising receiving a video frame, comparing said video frame to background reference frame to locate difference, locating a plurality of objects to form a plurality of marked objects; and determining a behavior pattern for an object according to the difference, said behavior pattern is defined according at least one scene characteristic. The method further comprises producing an updated background reference frame. The method further comprising determining the difference performed by creating a difference frame between the video frame and the background reference frame. The method further comprises finding a new object when determining the difference and an alarm according to said behavior pattern. A pre-defined pattern of suspicious behavior comprises an object presenting unpredictable behavior.

A second aspect of the present invention relates to a system for analyzing video data comprising a plurality of video frames, the system comprising, a video frame preprocessing layer for determining a difference between a plurality of video frames, an object clustering layer for detecting a plurality of objects according to said difference, and an application layer for characterizing said plurality of objects according to scene characteristic. The difference is determined between a video frame and a reference frame. The system further comprises a background refreshing layer for preparing an updated reference frame according to the said difference. The scene characteristic defines a behavior pattern for an object, such that if the object exhibits the behavior pattern, the scene characteristic is detected. If the scene characteristic is detected,

an alarm is generated. The scene characteristic further comprises a parameter for determining if the object exhibits the behavior pattern.

A third aspect of the present invention refers to a system for detecting a vehicle remaining in a restricted zone for at least a minimum period of time, comprising, a video content analysis module for analyzing video data of the restricted zone, said video content analysis module further comprising an object tracking component, and an application layer for receiving data from said video content analysis module and for detecting a vehicle remaining in the restricted zone for the minimum period of time, and said application layer generating an alarm upon detection.

A fourth aspect of the present invention refers to a system for detecting unattended luggage, bag or any unattended object in an area, comprising, a video content analysis module for analyzing video data of the area, said video content analysis module further comprising an object tracking component, and an application layer for receiving data from the video content analysis module and for detecting an unattended object, wherein said unattended object has not been attended in the area for more then a predefined period of time.

A fifth aspect of the present invention refers to a surveillance system for the detection of an alarm situation, the system comprising the elements of, a video analysis unit for analyzing video data representing images of a monitored area, the video analysis unit comprising an object tracking module to track the movements and the location of a video object, a detection, surveillance and alarm application for receiving video data analysis results from the video analysis unit, for identifying an alarm situation and to generate an alarm signal, an events database to hold video objects, video object parameters and events identified by the application. The system comprises the elements of, an application driver to control the detection, surveillance and alarm application, a database handler to access, to update and to read the events database, a user interface component to communicate with a user of the system, an application setup and control component to define the control parameters of the application, an application

setup parameters table to store the control parameters of the application. The system further comprises the elements of, a video data recording and compression unit to record and compress video data representing images of a monitored area, a video archive file to hold the recorded and compressed video data representing images of the a monitored area, an alarm distribution unit to distribute the alarm signal representing an alarm situation. The system further comprises the elements of, a video camera to obtain the images of a monitored area, a video capture component to capture video data representative of the images of the monitored area, a video transfer component to transfer the captured video data to the video analysis unit and the recording compressing and archiving unit, a computing and storage device. The object tracking module comprises the elements of, a video frame preprocessing layer for determining the difference between video frames, an objects clustering layer for detecting objects in accordance with the determined difference, a scene characterization layer for characterizing the object according to characteristic of a scene, a background refreshing layer for preparing an updated reference according to the determined difference. The detection surveillance and alarm application is operative in the detection of an unattended object in the monitored area. Any video camera within the system, the video capturing component, the video transfer component and the computing and storage device can be separated and can be located in different locations. The interface between the video camera, the video capturing component, the video transfer component and the computing and storage device is a local or wide area network or a packed-based or cellular or radio frequency or micro wave or satellite network. The unattended object is a luggage left in an airport terminal for a pre-determined period or a vehicle parking in a restricted zone for a pre-defined period. The detection surveillance and alarm application is operative in the detection of an unpredicted object movement. The analysis is also performed on audio data or thermal imaging data or radio frequency data associated with the video data or the video object in synchronization with the

video data. The video capture component captures audio or thermal information or radio frequency information in synchronization with the video data.

A sixth aspect of the present invention refers to a surveillance method for the detection of an alarm situation, the surveillance to be performed on a monitored scene having a camera, the method comprising the steps of, obtaining video data from the camera representing images of a monitored scene, analyzing the obtained video data representing images of the object within the monitored scene, the analyzing step comprising of identifying the object within the video data, and inserting the identified object and the event into an event database. Another embodiment of the method further comprises the steps of, retrieving of the object associated with an event, according to user instruction displaying the video event associated with the retrieved object. The method further comprises the steps of, retrieving at least two events, associating according to parameters of the object, the object with the at least two events. The method comprises the steps of, debriefing the object associated with the event to identify the pattern of behavior or movement of the object within the scene within a predefined period of time. The method further comprises the steps of, pre-defining patterns of suspicious behavior; and pre-defining control parameters. The method further comprises the steps of, recognizing an alarm situation according to the pre-defined patterns of suspicious behavior, and generating an alarm signal associated with the recognized alarm situation. The method further comprises the steps of, implementing patterns of suspicious behavior introducing pre-defined control parameters, recording, compressing and archiving the obtained video data, distributing the alarm signal representing an alarm situation across a pre-defined range of user devices. The pre-defined pattern of suspicious behavior comprises, an object entering a monitored scene, the object separating into a first distinct object and a second distinct object in the monitored scene, the first distinct object remaining in the monitored scene without movement for a pre-defined period, and the second distinct object leaving the monitored scene. The pre-defined pattern of suspicious behavior comprises, an object entering the

monitored scene, the object ceasing its movement, the size of the object is recognized as being above a pre-defined parameter value, and the object remaining immobile for a period recognized as being above a pre-defined parameter value. The method does further comprise identifying information associated with the object for the purpose of identifying objects.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the drawings in which:

Fig. 1 is a schematic block diagram of the system architecture, in accordance with the preferred embodiments of the present invention;

Fig. 2 is a simplified flowchart that illustrates the operation of the object tracking method, in accordance with the preferred embodiment of the present invention;

Fig. 3A is a simplified flowchart illustrative of the operation of unattended luggage detection application, in accordance with the first preferred embodiment of the present invention;

Fig. 3B illustrates the control parameters of the unattended luggage detection application, in accordance with the first preferred embodiment of the present invention;

Fig. 4A is a simplified flowchart illustrative of the operation of the city center application, in accordance with the second preferred embodiment of the present invention;

Fig. 4B illustrates the control parameters of the city center application, in accordance with the second preferred embodiment of the present invention; and

Fig. 5 is a flowchart describing the operation of the proposed method, in accordance with the preferred embodiments of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

A monitoring and surveillance system and method for the detection of potential alarm situation via a recorded surveillance content analysis and for the management of the detected unattended object situation via an alarm distribution mechanism is disclosed. The proposed system and method includes an advanced architecture and a novel technology operative in capturing surveillance content, analyzing the captured content and providing in real time a set of alarm messages to a set of diverse devices. The analysis of the captured content comprises a unique algorithm to detect, to count and to track objects embedded in the captured content. The present invention provides a detailed description of the applications of this method. The method and system of the present invention may be implemented in the context of unattended objects (such as luggage, vehicles or persons), parking or driving in restricted zones, controlling access of persons into restricted zones, preventing loss of objects such as luggage or persons and counting of persons.

In the preferred embodiments of the present invention, the monitored content is a video stream recorded by video cameras, captured and sampled by a video capture device and transferred to a video processing unit. The video processing unit performs a content analysis of the video images and indicates an alarm situation in accordance with the results of the analysis. In other preferred embodiment of the invention, diverse other content formats are also analyzed, such as thermal based sensor cameras, audio, wireless linked camera, data produced from motion detectors, and the like.

The first preferred embodiment of the present invention concerns the detection of unattended objects, such as luggage in a dynamic object-rich environment, such as an airport or city center. The second preferred embodiment of the invention concerns the detection of a vehicle parked in a forbidden zone, or the extended-period presence of a non-moving vehicle in a restricted-period parking zone. Forbidden or restricted parking zones are typically associated with

sensitive traffic-intensive locations, such as a city center. Another preferred embodiment of the invention concerns the tracking of objects such as persons in various scenarios, such as a person leaving the vehicle away from the terminal, which may equal suspicious (unpredicted) behavioral pattern. In other possible embodiments of the present invention the system and method can be implemented to assist in locating lost luggage and to restrict access of persons or vehicles to certain zones. Other preferred embodiments of the invention could regard the detection of diverse other objects in diverse other environments. The following description is not meant to be limiting and the scope of the invention is defined only by the attached claims.

Referring to Fig. 1 a set of video cameras 12, 14, 16, 18 operate in a security-wise sensitive environment and cover a specific pre-defined zone that is required to be monitored. The area monitored can be any area preferably in a transportation area including an airport, a city center, a building, and restricted or non-restricted areas within buildings or outdoors. The cameras 12, 14, 16, 18 could be analog devices or digital devices. The cameras can capture normal light, infra-red, temperature, or any other form of radiation. By using audio capturing devices such as a microphone (not shown), the cameras can also capture auditory signals, such as noise generated by machines and voices made by persons. The cameras 12, 14, 16, 18 continuously acquire and transmit sequences of video images to a display device 21, such as a video terminal, operated by a human operator. The display device 21 could be optionally provided with video images from the vide archives 40 by the computing and storage device 24. The cameras, 12, 14, 16, 18 transmit sequences of video images to a video capture component 20 via suitably wired connections. The video capture component 20 could capture the images through an analog interface, a digital interface or through a Local Area Network (LAN) interface or Wide Area Network (WAN), IP, Wireless, Satellite connectivity. The video capture component 20 can be a NICEVISION system manufactured by Nice Systems Ltd., Raanana, Israel. The video capture

component 20 can also be configured to capture audio signals captured by the cameras audio capturing devices. The audio and video signals are preferably synchronized at the system level. The component 20 receives the sequences of video images and appropriately samples the video stream. Where the processing of the captured video stream is performed by an external computing platform, such as a Personal Computer (PC), a UNIX workstation, or a mainframe computer, the unit 20 sends the sampled video information to a video transfer component 22. The video transfer component 22 transfers the video information to a computing and storage device 24. The device 24 could be an external computing platform, such as a personal computer (PC), a UNIX workstation or a mainframe computer having appropriate processing and storage units or a dedicated hardware such as a DSP based platform. It is contemplated that future hand held devices will be powerful enough to also implement device 24 there within. The device 24 could be also an array of integrated circuits with built-in digital signal processing (DSP) and storage capabilities attached directly to the video capture component 20. The capture component 20 and the transfer component 22 are preferably separate due to the fact that a capture component can be located at the monitored scene, while a transfer component can be located away from the monitored scene. In another preferred embodiment the capture and the transfer components can be located in the same device. The device 24 includes a video analysis unit 26, an application driver 30, a database handler 32, a user interface 34, a setup and control Detection Surveillance and Alarm (DSA) application 36, a recording compressing and archiving unit 38, a video archives 40, an events database 42, an alarm distribution unit 44, a DSA application and setup parameters file 46, and DSA application 48. Optionally, whenever video is captured and processed, audio signals captured in association with these captured video signals can be stored and tagged as relating to the video captured and processed.

Still referring to Fig. 1 the video content is transferred optionally to the recording, compressing, and archiving unit 38. The unit 38 optionally compresses

the video content and stores the compressed content to the video archive files 40. The video archive files 40 could be suitable auxiliary storage devices, such as optical disks, magnetic disks, magnetic tapes, or the like. The stored content is held on the file 40 for a pre-defined (typically long) period of time in order to enable re-play, historical analysis, and the like. In parallel the video content is transferred to the video analysis unit 26. The unit 26 receives the video input, activates the object tracking module 28 and activates the application driver 30. In accordance, with results of the video analysis performed in conjunction with the object tracking module 28, the video analysis unit 26 further generates the appropriate alarm or indication signal where a specific alarm situation is detected. The application driver 30 includes the logic module of the application 48. The driver 30 receives event data and alarm data from the video analysis unit 26 and inserts the event data and the alarm data via the database handler 32 into the events database 42. The driver 30 further controls the operation of the DSA application 48. The setup and control DSA application 36 is used by the user of the system in order perform system setup, to define control parameters, and the like. The user interface 34 is responsible for the communication with the user. The event database 42 stores the event data and the alarm data generated by the video analysis unit 26. The event database 42 also holds the search parameters for searching objects or events for the purpose of investigating the events or objects. The search parameters include the object circle-like shape and object location parameters. Other object search parameters can also include data collected from various cameras, which may have captured the same object. The collected data could provide important information about the object, such as object type (animate or inanimate), object identification (via face recognition), color histogram (color of hair, of cloth, of shoes, of complexion), and the like. The parameters allow finding associations between objects and events captured by different cameras. The alarm distribution unit 44 optionally distributes the received alarm signals to a variety of alarm and messaging device. The DSA application and setup parameter file 46 stores the setup information and

parameters generated by the user via the setup and control DSA application 36. The DSA application 48 provides real-time video to the user, performs re-plays of video by request, submits queries to the event database 42 and provides alarm messages, such as suitably structured pop-up windows, to the user via the user interface 34.

Still referring to Fig. 1 the units and components described could be installed in distinct devices distributed randomly across a Local Area Network (LAN) that could communicate over the LAN infrastructure or across Wide Area Networks (WAN). One example is a Radio Frequency Camera that transmits composite video remotely to a receiving station, the receiving station can be connected to other components of the system via a network or directly. The units and components described could be installed in distinct devices distributed randomly across very wide area networks such as the Internet. Various means of communication between the constituent parts of the system can be used. Such can be a data communication network, which can be connected via landlines or cellular or like communication devices and that can be implemented via TCP/IP protocols and like protocols. Other protocols and methods of communications, such as cellular, satellite, low band, and high band communications networks and devices will readily be useful in the implementation of the present invention. The components could be further co-located on the same computing platform or distributed across several platforms for load balancing. The components could be redundantly replicated across several computing platforms for specific operational purposes, such as being used as back-up systems in the case of equipment failure, and the like. Although on the drawing under discussion only a limited set of cameras and only a single computing and storage device are shown it will be readily perceived that in a realistic environment a plurality of cameras could be connected to a plurality of computing and storage devices.

Referring now to Fig. 2 showing illustrates the operation of the object tracking method. The proposed system and method is based upon a video content analysis method that can detect, track and count objects in real time in accordance

with the results of the video stream processing. This method enables the detection of new objects created from the object identified and tracked. The method also enables to identify when an object tracked has merged with another object. The merging of objects with other objects and the creation of objects from other objects is particularly important to identify persons leaving vehicles, luggage, or other objects in the environment monitored. The ability to detect if an object is created or disappears also enables the method of the present invention to identify if persons disturb objects or move objects. The method is implemented in the object tracking module 28 of Fig. 1. The method receives the following input: new video frame, background (reference frame), detected objects from the last iteration. The outputs of the method comprise the updated background, and the updated objects. Fig. 2 illustrates the four layers that jointly implement object tracking and detection method. The video frame pre-processing layer 52 uses a new frame and one or more reference frame for generating a difference frame representing the difference between the new frame and the reference frame or frames. The reference frame can be obtained from one of the capture devices described in association with Fig. 1 or provided by the user. The difference frame can be filtered or smoothened. The objects clustering layer 54 generates new/updated objects from the difference frame and the last known objects. The scene characterization layer 56 uses the objects from the objects clustering layer 54 in order to describe the scene. The background-refreshing layer 58 updates the background (reference frame or frames) for the next frame calculation and a refreshing process uses the outputs of all the previous layers to generate a new reference layer or layers. Note should be taken that in other preferred embodiments of the invention other similar or different processes could be used to accomplish the underlying objectives of the system and method proposed by the present invention.

The first preferred embodiment of the invention regards an unattended object detection system and method. The unattended object could be a suitcase, a carrier bag, a backpack, or any other object that was left unattended in a security-

sensitive area, such as an airport terminal, a train station's waiting room, a public building, or the like.

Referring now to Fig. 3A showing a flowchart illustrative of the operation of unattended luggage detection application of the method of the present invention. An unattended luggage event is detected via the performance of a sequence of operative steps. The logic underlying the performance is based on a specific scenario. In the scenario it is assumed that a terrorist or any other individual having criminal intent may enter a security-sensitive area carrying a suitcase. In one example, the suitcase may contain a concealed explosive device. In another example, the suitcase may have been lost or unattended for a prolonged length of time. In yet another application, a suitcase or an object may have been taken without authority. In the first example, the individual may surreptitiously (in a manner that non-recognizable by the monitoring cameras) activate a time-delay fuse mechanism connected to the explosive device that is operative in the timed detonation of the device. Subsequently, the individual may abandon the suitcase unattended and leave the security-sensitive area for his own safety in order not to be exposed to the damage effected by the expected detonation of the explosive device. The following operative conclusions, indicating certain sub-events, are reached by the suitable execution of sets of computer instructions embedded within a specifically developed computer program. The program is operative in the analysis of a sequence of video images received from a video camera covering a security-sensitive area, referred herein below to as the video scene. An unattended luggage event is identified by the program when the following sequence of sub-events takes place and detected by the execution of the program: a) an object enters the video scene (62). It is assumed that the object is a combined object comprising an individual and a suitcase where the individual carries the suitcase. b) The combined object is separated into a first separate object and a second separate object (64). It is assumed that the individual (second object) leaves the suitcase (first object) on the floor, a bench, or the like. c) The first object remains in the video scene

without movement for a pre-defined period of time (66). It is assumed that the suitcase (first object) was left unattended. d) The second object exits the video scene (68). It is assumed that the individual (second object) left the video scene without the suitcase (first object) and is now about leave the wider area around the video scene. Following the identification of the previous sub-events, referred to collectively as the video scene characteristics, the event will be identified by the system as a situation in which an unattended suitcase was left in the security-sensitive area. Thus, the unattended suitcase will be considered as a suspicious object. Consequently, the proposed system may generate, display and/or distribute an alarm indication. Likewise, in an alternative embodiment in step 62, a first object, such as a suitcase or person monitored is already present and monitored within the video scene. Such object can be lost luggage located within the airport. Such object can be a person monitored. In step 64 the object merges into a second object. The second object can be a person picking up the luggage, another person to whom the first person joins or a vehicle to which the first person enters. In step 66 the first object (now merged with the second object) moves from its original position and at step 68 of the alternative embodiment exists the scene. The system of the present invention will provide an indication to a human operator. The indication may be oral, visual or written. The indication may be provided visually to a screen or delivered via communication networks to officers located at the scene or to off-premises or via dry contact to an external device such as a siren, a bell, a flashing or revolving light and the like.

Referring now to Fig. 3A which illustrates the control parameters of the unattended luggage detection application. In order to set up the unattended luggage detection application, the user is provided with the capability of defining the following control parameters: a) area or areas within the scanned zone wherein the system will search for suspected objects (70), b) the dimensional limits of the detected object (72). Objects having dimensions out of the limits defined will not be detected as suspected objects, and c) a time out value that is the amount of time that should pass from the point-in-time at which the suspected

object was detected as non-moving and the point-in-time until an alarm is generated. Once an alarm is raised the officer reviewing the monitored scene may request the system to provide a playback so as to identify the objects in question. Once playback resumes the officer may tap on a touch sensitive screen (or select the image by other means such as a mouse, a keyboard, a light pen and the like) and the system may play back the history of video captured in association with the relevant object or objects. If a second object, such as a luggage left unattended is played back, the playback will identify the person taking or leaving the object. The officer may select such person and request play back or forward play to ascertain where the person came from or where the person went to and appropriately alert security officers. In order to debrief the event the officer may mine the database in various ways. One example would be to request the system to retrieve the events or objects that are similar to search parameters associated with the object or event he is investigating that can help in identifying the location of person or objects or the whereabouts or actions performed by the object. A suspect may place a suitcase given to him previously (not in the same scene) by a third person and leave the airport in a vehicle. Once the officer is alerted to the fact that the suitcase is unattended he may investigate the retrieve the third party associated with the handing of the suitcase and the vehicle associated with the suspect. The system, in real time, stores from each camera the various objects viewed. In the ordinary course of events the system would associate between like objects captured by various cameras using the initial search parameter (such as the suspect's parameters). During the investigation stage the system retrieves the associated objects and the event linked therewith and presents the events and object to the viewing officer in accordance with his instructions. The officer may decide to review forward or backward in time scenes the system would mark the associated objects thus allowing the officer to identify the stream of events elected for a particular object, such as existing a vehicle, handing over a suitcase, leaving a suitcase unattended, walking in unpredicted directions and the like.

As noted above the user may provide a predefined background. The background may be captured from the capturing devices. The human operator may define elements within the screen as background elements. Such can be moving shades or areas of little interest and the like.

Note should be taken that the above-described steps for the detection of a suspected object and the associated control parameters are exemplary only. Diverse other sequences of steps and different control parameters could be used in order to achieve the inherent objectives of the present invention.

The second preferred embodiment of the invention regards a detection of vehicles parked in restricted area or moving in restricted lanes. Airports, government buildings, hotels and other institutions typically forbid vehicles from parking in specific areas or driving in restricted lanes. In some areas parking is forbidden all the time while in other areas parking is allowed for a short period, such as several minutes. In the second preferred embodiment of the invention a system and method is proposed that detect vehicles parking in restricted areas for more than a pre-defined number of time units and generates an alarm when identifying an illegal parking event of a specific vehicle. In another preferred embodiment the system and method of the present invention can detect whether persons disembark or embark a vehicle in predefined restricted zones. The use of the embodiment described in association with Figs. 3A, 3B can be employed in association with the application of the invention described below in association with Figs. 4A, 4B.

Referring now to Fig. 4A which shows an exemplary flowchart illustrative of the operation of the city center application. An illegal parking event is detected via the performance of a sequence of operative steps. The logic underlying the performance is based on a specific exemplary scenario. In the scenario it is assumed that one or more persons may drive into a security-sensitive area in a specific vehicle. The vehicle could contain a powerful hidden explosive device that designed to be activated by one of the occupants of the vehicle or persons later embarking said vehicle in a restricted zone. Alternatively, the one or

more occupants of the vehicle could be armed and may plan an armed attack against specific targets, such as individuals entering or exiting the building. In the first case scenario one of the occupants may surreptitiously (in a manner that is non-recognizable by the monitoring cameras) activate a time-delay fuse mechanism connected to the explosive device that is operative in the timed detonation of the device. Subsequently, the occupants may abandon the vehicle and leave the security-sensitive area for their own safety in order not to be exposed to the damage effected by the expected detonation of the explosive device. In a second case scenario, the occupants may remain in the vehicle while waiting for the potential target, such as an individual about to enter the scene either from the building or driving into the area. In another embodiment the vehicle may park in a restricted zone where such parking is not allowed in specific hours or where parking or standing is restricted for a short period of time. The following operative conclusions are achieved by the suitable execution of sets of computer instructions embedded within a specifically developed computer program. The program is operative in the analysis of a sequence of video images or other captured data received from the cameras covering a security-sensitive area (referred herein under to as the video scene). An illegal parking event is identified by the program when the following sequence of sub-events takes place and detected by the execution of the program: a) an object enters the video scene (76). The system identifies the object and in accordance with size and shape is assumed to be a vehicle occupied by one or more individuals. b) The object has subsequently to entering the restricted zone stopped moving (78). It is assumed that the vehicle has stopped. c) The dimensions of the object are above a pre-defined dimension parameter value (80). The dimensions of the object are checked in order to distinguish between vehicles and pedestrians that may enter the same area and stop therein (for example, sitting down on a bench). The size of the object is also important to determine the direction of movement of the object. Thus, objects growing bigger are moving in the direction of the capturing device while objects whose size is reduced over time are assumed to be moving away

from the capturing device. Certain predefined parameters as to size may be pre-programmed into the system in association with a specific background image. d) The object stays without motion for a pre-defined period (82). As described in association with Figs. 3A, 3B the system may detect whether an occupant of the object has left the object and is now about to leave the wider area around the video scene. Following the identification of the previous sub-events, referred to as specific video scene characteristics, the event will be identified by the system as a situation in which an illegally parked vehicle was left in the security-sensitive area unattended. Thus, the parked vehicle will be considered as a suspicious object. Another event, which is recognized as suspicious, is where the vehicle is moving in an unpredicted direction or if an object, such as a person, is leaving the vehicle and moving to an unpredictable direction. An predicted direction can be predefined and an unpredicted direction is the direction opposite or a direction which does not match a predefined direction of flow of persons or vehicles. For example, in an airport the sidewalk where persons disembark from vehicles can be defined as a predicted direction and the side opposite the sidewalk and across the lanes of travel can be defined as the unpredicted direction. Thus, if persons disembarking vehicles at the airport leave the vehicle unattended (stationary for a predefined period of time) and walk or run not towards the airport, but rather to the opposite the system identifies a suspicious event. In city centers the same can be applied near bus depots or train stations and also in retail establishments monitoring areas not for public access.

Another parameter, which can be viewed, is the speed of the object. Speeding away from the vehicle can be an additional indicator that a suspicious event is taking place. The parked vehicle may also be regarded as suspicious if it is parked in the restricted zone more than a predefined period of time. Consequently, the proposed system may generate, display and/or distribute an alarm indication. Alternatively, if the occupants of the vehicle did not leave the vehicle but still wait in the vehicle an alert can be raised, assuming a person is

waiting in the vehicle in suspicious circumstances or parking illegally. In this scenario too the parked vehicle will be considered a suspicious object. Consequently, the proposed system may generate, display and/or distribute an alarm indication. Once an alarm is raised the officer reviewing the monitored scene may request the system to provide a playback so as to identify the objects in question. Once playback resumes the officer may tap on a touch sensitive screen (or select the image by other means such as a mouse, a keyboard, a light pen and the like) and the system may play back the history of video captured in association with the relevant object or objects. If a second object, such as a person disembarked the vehicle the officer may tap the object and request a follow up playback associated just with that person. The playback or play forward feature allows the officer to make a real time determination as to the objects nature including information stored in the database (such as parameter association with the object) and determine the next action to be taken.

In another embodiment an alert may be raised as soon as an object in the size of a vehicle as determined by the relative size of the object as predefined in the system enters a restricted lane. The application concerning restricted lanes may check the size of the vehicles in such lanes as bus lanes wherein only buses (which are larger than vehicles) are allowed. If the object is a vehicle, i.e. smaller than a bus, an alert may be raised. The system may identify the vehicle and later a ticket may be issued to the owner of the vehicle. This application is extremely useful for policing restricted lanes without having a police unit on the scene.

In another embodiment of the invention, a database of recognized vehicle plate numbers can be utilized to assist in the off line investigation and associated identification of the owner of a suspicious vehicle. The database can also be used to determine whether the number of the license plate is stolen or belongs to a suspect on a pre-supplied list.

Referring now to Fig. 4B the exemplary control parameters of the city center application are shown. In order to set up the illegal parking event detection application, the user is provided with the capability of defining the following

parameters: a) area or areas within the scanned zone wherein the system will search for suspected objects (84), b) the dimensional limits of the detected object (86). The minimal dimensional values provided in order to limit the type of object as a vehicle, and c) a time out value (88) that is the amount of time that may pass between the point-in-time at which the object stopped moving and the point-in-time where an alarm will be generated.

Note should be taken that the above-described steps for the detection of an illegally parked vehicle and the associated control parameters are exemplary only. Diverse other sequences of steps and different parameters could be used in order to achieve the inherent objectives of the present invention.

Referring now to Fig. 5 which is a flowchart describing the operation of the proposed method. The fundamental logical flow of the method is substantially similar for both the first preferred embodiment concerning the unattended luggage detection and for the second preferred embodiment concerning the detection of an illegally parked vehicle. At step 93 a set of suspicious behavior patterns are defined. At step 95 a set of application and control parameters are defined. Both steps 93 and 95 are performed offline. The behavior patterns are implemented in the DSA application 46 of Fig. 1. while the control parameters are stored in the application and setup parameter file 46 of Fig. 1. At step 92 video data of a pre-defined video scene is acquired by one or more video cameras. At step 94 the video stream is transmitted to the video capture unit and at step 96 the video data is sampled. Optionally the video data is recorded, compressed and archived on auxiliary storage devices, such as disks and/or magnetic tapes (step 98). Simultaneously the video input is transferred to the video analysis component. The object detection unit analyzes the video data (step 100) and activates a specific DSA application 100. In the first preferred embodiment the DSA application 100 is the unattended luggage detection while in the second preferred embodiment the DSA application 100 is the detection of an illegally parked vehicle. Another DSA application 100 is a lost object prevention application. Other applications are evident from the description

provided above in association with Figs. 3A, 3B, 4A, and 4B. At step 104 it is determined whether, in accordance with the specific application and the associated parameters, including optional database parameters, an alarm state was detected by the video analysis unit. If no alarm state was identified then program control returns to the application 102. When an alarm state was raised at step 106 the alarm event is inserted into the event database, at step 108 a suitable message including alarm state details is sent to Graphic User interface (GUI) control application, at step 110 an optional message including alarm state-specific details is sent to the optional alarm distribution system to be distributed to a set of pre-defined monitor devices, such as wireless, personal data assistance devices, pagers, telephones, e-mail and the like.

The GUI control application 108 prompts the user for a suitable response concerning the alarm or optionally presents the user in real time with the video data sent by the camera the output of which generated the alarm. The alarm can be provided as text or pop up window on the screen of the operator, as e-mail sent to an officer, SMS message sent to a cellular phone, an automated telephone call to an officer, a text pager message, pictures or video stream sent the officer's portable device or hand held device, or send via a dry contact to generate a siren or an audio or visual indication and the like. The message could be provided to one or many persons or to specific persons associated with the specific event or alarm. The suspicious object on the video images is emphasized in a graphic manner, such as encircling the object in a circle-like or oval graphic element that is overlaid on the video image. Other information concerning the object, such as the object its size, speed, direction of movement, range from camera, if identified and the like, will appear next to the object's image or in another location on the screen. If the optional recording and archiving unit and the associated video archive files are implemented on the system then the user is provided with the option of video data re-play. When the optional alarm distribution component is implemented on the system, the alarm message will be appropriately distributed to a set of pre-defined and suitable pre-configured locations.

Objects monitored by surveillance systems may move in unpredicted directions. In For example, in an airport surveillance scene a person may arrive to the scene with a suitcase, enter the terminal building, leave the suitcase near the entrance of a terminal, and then leave the terminal. In another similar example, a first object (a vehicle) may arrive to the entrance of a terminal, a person (second object) may exit the vehicle, walk away in a direction opposite to the terminal, thus leaving the scene. In order to recognize patterns of unpredictable behavior a set of pre-defined rules could be implemented. These rules assist the system in capturing unpredictable behavior patterns taking place within the scenes monitored by the system.

The present system collects and saves additional information relating to each object. An initial analysis is performed in connection with each object. Apart from the circle-like shape and location of the object, the system attempts to identify whether the object is a person or an inanimate object. In addition, the object will collect and save object parameters such as the object-normalized size, distance from camera, color histogram. If the person is a person a face recognition algorithm is activated to try and determine whether the person is recognized. Recognized persons can be those persons that have been previously identified in other objects or may be faces that are provided to the system, such as from law enforcement agencies or that are previously scanned by the employer. Other parameters may also be associated with the object such as name, other capturing devices, speed and the like.

When sufficient computing power is available, the system would also perform in real time a suitable analysis of the object in order to create associated search parameters, such as, for example, color histogram and other search parameters mentioned above and to immediately alert officers if the analysis leads to predetermined alarm status, such as when the a particular face is recognized which is a wanted person or a person not allowed or recognized in a restricted zone. In addition, in on line mode the system can identify more than one

parameters, such as a non-recognized face in a restricted zone and speaking in a foreign language, or a person not wearing a particular identifying mark (such as a hat or a shirt in a particular color) and the person is exiting a vehicle.

The proposed system and method provide real-time and off-line processing of suspicious events. For example, when a vehicle arrives at a terminal of an airport or train station and a person leaves the vehicle to a direction opposite the terminal, the present system and method will automatically alert the user. Such suspicious behavioral patterns are predetermined and the present system and method analyzes events to detect such events. The present system and method is further capable of identifying a set of linked events associated with the same object. An object can be defined as any detected object that continues to move within the captured scene. An event is defined as a series of frames capturing a scene and objects there within. The event can be associated with a particular capturing device. Linked events to the same object relate to a single object in the same area throughout the surveillance period whether captured by one or more cameras and appearing in one or more events. The system will track (either upon request or automatically) an object through one or more events. The present system and method also provide the ability to associate a retrieved event or object with unique parameters of such an object, in addition to the object oval characteristics and location. Such would include, for example face recognition, color of clothes through the use of a histogram color. The difference between the color of the clothes and the color of the shirt of an object, color per zone in the object, such as the color of a hair, normalized size subject to the distance from the camera, and normalized shape of objects such as the size of a suitcase. The use of object associated parameters in addition to the object's shape and position enable the post event data base search of an object according to the parameters to quickly obtain the event or events associated with the object or other objects associated with the object. Such parameters also enable the user of the present invention to investigate and request the system to identify a particular object or event. This enables a better retrieval of the events and objects. The system may also, in real

time, associate the parameters with objects and perform rule checking to determine if the objects comply with rules that are permitted in the scene, such as objects are not left unattended, objects move is specific directions, objects do not depart from other objects in specific locations, and the like.

The additional embodiments of the present system and method will now be readily apparent to person skilled in the art. Such can include crowd control, people counting, an offline and online investigation tools based on the events stored in the database, assisting in locating lost luggage (lost prevention) and restricting access of persons or vehicles to certain zones. The applications are both for city centers, airports, secure locations, hospitals and the like.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined only by the claims, which follow.

CLAIMS

WHAT IS CLAIMED IS:

1. A method for analyzing video data, comprising:
receiving a video frame;
comparing said video frame to at least one background reference frame to locate at least one difference;
locating a plurality of objects to form a plurality of marked objects; and
determining a behavior pattern for at least one object according to said at least one difference, wherein said behavior pattern is defined according at least one scene characteristic.

2. The method of claim 1, further comprising producing at least one updated background reference frame.

3. The method of claim 1, wherein determining said at least one difference is performed by creating a difference frame between said video frame and said at least one background reference frame.

4. The method of claim 1, further comprising finding at least one new object when determining said at least one difference.

5. The method of claim 1, further comprising issuing an alarm according to said behavior pattern.

6. The method as claimed in claim 1 wherein a pre-defined pattern of suspicious behavior comprises an object presenting unpredictable behavior.

7. A system for analyzing video data comprising a plurality of video frames, the system comprising:

a video frame preprocessing layer for determining a difference between a plurality of video frames;

an object clustering layer for detecting a plurality of objects according to said difference; and

an application layer for characterizing said plurality of objects according to at least one scene characteristic.

8. The system of claim 7, wherein said difference is determined between a video frame and at least one reference frame, the system further comprising a background refreshing layer for preparing at least one updated reference frame according to the said difference.

9. The system of claim 8, wherein said at least one scene characteristic defines a behavior pattern for at least one object, such that if said at least one object exhibits said behavior pattern, said at least one scene characteristic is detected.

10. The system of claim 9, wherein if said at least one scene characteristic is detected, an alarm is generated.

11. The system of claim 9, wherein said at least one scene characteristic further comprises at least one parameter for

determining if said at least one object exhibits said behavior pattern.

12. A system for detecting a vehicle remaining in a restricted zone for at least a minimum period of time, comprising:

a video content analysis module for analyzing video data of the restricted zone, said video content analysis module further comprising an object tracking component; and

an application layer for receiving data from said video content analysis module and for detecting a vehicle remaining in the restricted zone for at least the minimum period of time, and said application layer generating an alarm upon detection.

13. A system for detecting unattended luggage, bag or any unattended object in an area, comprising:

a video content analysis module for analyzing video data of the area, said video content analysis module further comprising an object tracking component; and an application layer for receiving data from said video content analysis module and for detecting an unattended object, wherein said unattended object has not been attended in the area for more then a predefined period of time.

14. A surveillance system for the detection of an alarm situation, the system comprising the elements of:

a video analysis unit for analyzing video data representing images of a monitored area, the video analysis unit

comprising an object tracking module to track the movements and the location of a video object;

a detection, surveillance and alarm application for receiving video data analysis results from the video analysis unit, for identifying an alarm situation and to generate an alarm signal;

an events database to hold video objects, video object parameters and events identified by the application.

15. The system of claim 14 further comprises the elements of:

an application driver to control the detection, surveillance and alarm application;

a database handler to access, to update and to read the events database;

a user interface component to communicate with a user of the system;

an application setup and control component to define the control parameters of the application;

an application setup parameters table to store the control parameters of the application.

16. The system as claimed in claim 14 further comprises the elements of:

a video data recording and compression unit to record and compress video data representing images of a monitored area;

a video archive file to hold the recorded and compressed video data representing images of the a monitored area;

an alarm distribution unit to distribute the alarm signal representing an alarm situation.

17. The system as claimed in claim 14 further comprises the elements of:

at least one video camera to obtain the images of a monitored area;

at least one a video capture component to capture video data representative of the images of the monitored area;

at least one video transfer component to transfer the captured video data to the video analysis unit and the recording compressing and archiving unit;

at least one computing and storage device.

18. The system as claimed in claim 14 wherein the object tracking module comprises the elements of:

a video frame preprocessing layer for determining the difference between at least two video frames;

an objects clustering layer for detecting at least one objects in accordance with the determined difference;

a scene characterization layer for characterizing the a least one object according to least one characteristic of a scene;

a background refreshing layer for preparing at least one updated reference according to the determined difference.

19. The system as claimed in claim 14 wherein the detection surveillance and alarm application is operative in the detection of at least one unattended object in the monitored area.

20. The system as claimed in claim 17 wherein any of the at least one video camera; the at least one video capturing component; the at least one video transfer component and the at least one computing and storage device can be separated and can be located in different locations.

21. The system as claimed in claim 17 wherein the interface between the at least one video camera; the at least one video capturing component; the at least one video transfer component and the at least one computing and storage device is a local or wide area network or a packed-based or cellular or radio frequency or micro wave or satellite network.

22. The system as claimed in claim 19 wherein the at least one unattended object is a luggage left in an airport terminal for a pre-determined period.

23. The system as claimed in claim 22 the at least one unattended object is a vehicle parking in a restricted zone for a pre-defined period.

24. The system as claimed in claim 14 wherein the detection surveillance and alarm application is operative in the detection of an unpredicted object movement.

25. The system as claimed in claim 14 wherein the analysis is also performed on audio data or thermal imaging data or

radio frequency data associated with the video data or the video object in synchronization with the video data.

26. The system as claimed in claim 17 wherein the video capture component captures audio or thermal information or radio frequency information in synchronization with the video data.

27. A surveillance method for the detection of an alarm situation, the surveillance to be performed on at least one monitored scene having at least one camera, the method comprising the steps of:

obtaining video data from the at least one camera representing images of a the at least one monitored scene;

analyzing the obtained video data representing images of the at least one object within the at least one monitored scene, the analyzing step comprising of identifying the at least one object within the video data; and

inserting the identified at least one object and the at least one event into an event database.

28. The method of claim 27 further comprising the steps of:

retrieving at least one of the object associated with at least one event;

according to user instruction displaying the video event associated with the retrieved at least one object.

29. The method of claim 27 further comprising the steps of:

retrieving at least two events;

associating according to parameters of the at least one object, the at least one object with the at least two events.

30. The method of claim 27 further comprising the steps of:
debriefing the at least one object associated with the at least one event to identify the pattern of behavior or movement of the at least one object within the at least one scene within a predefined period of time.

31. The method of claim 27 further comprising the steps of:
pre-defining patterns of suspicious behavior; and
pre-defining control parameters.

32. The method of claim 27 further comprising the steps of:
recognizing an alarm situation according to the pre-defined patterns of suspicious behavior; and
generating an alarm signal associated with the recognized alarm situation.

33. The method as claimed in 27 further comprises the steps of:
implementing patterns of suspicious behavior
introducing pre-defined control parameters;
recording, compressing and archiving the obtained video data;
distributing the alarm signal representing an alarm situation across a pre-defined range of user devices.

34. The method as claimed in claim 27 wherein a pre-defined pattern of suspicious behavior comprises:
an object entering a monitored scene;

the object separating into a first distinct object and a second distinct object in the monitored scene;

the first distinct object remaining in the monitored scene without movement for a pre-defined period; and

the second distinct object leaving the monitored scene.

35. The method of claimed in claim 34 wherein the pre-defined pattern of suspicious behavior comprises:

an object entering the monitored scene;

the object ceasing its movement;

the size of the object is recognized as being above a pre-defined parameter value; and

the object remaining immobile for a period recognized as being above a pre-defined parameter value.

36. The method of claim 35 further comprising identifying information associated with the object for the purpose of identifying the at least one object.

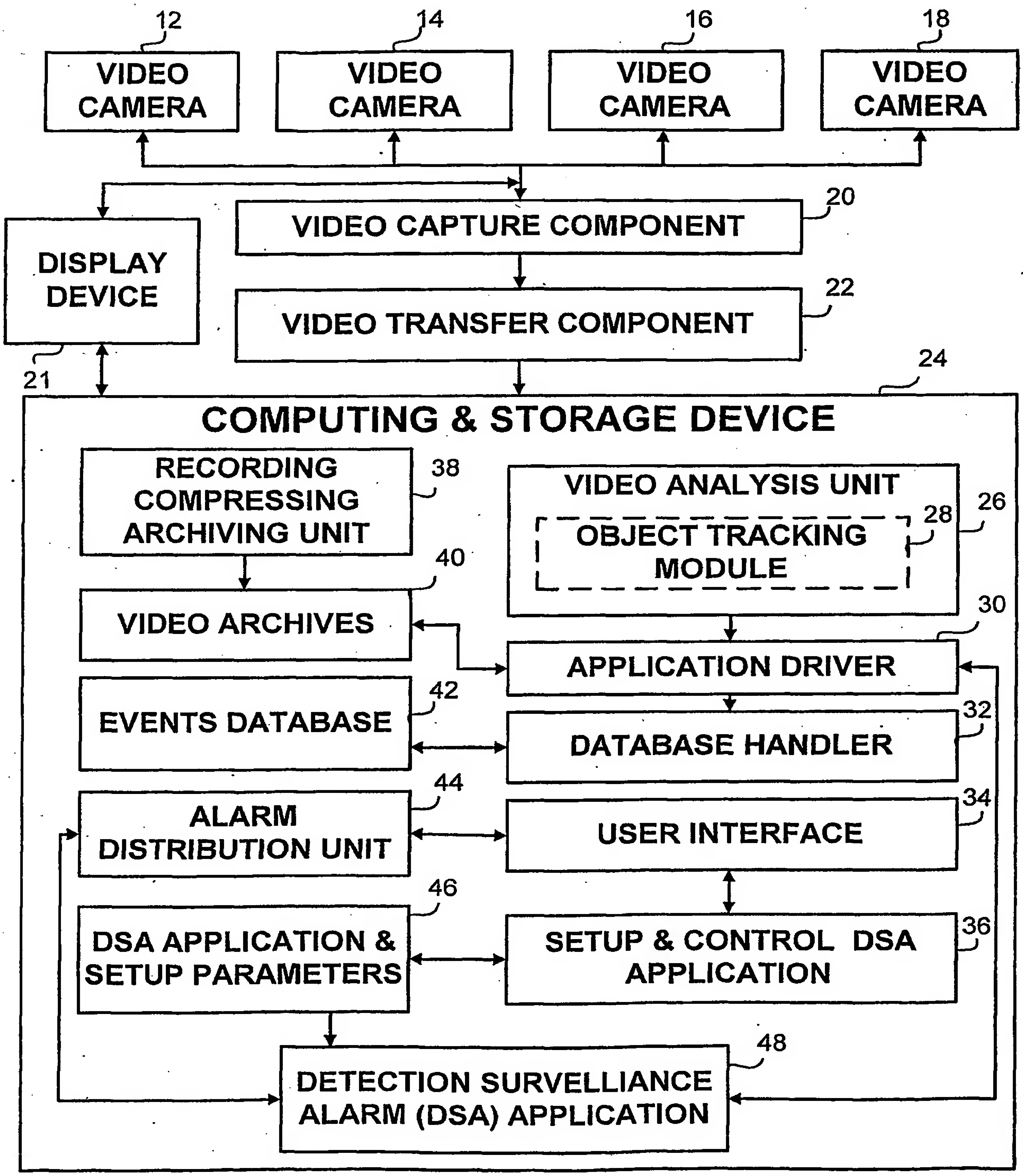

12
VIDEO CAMERA
14
VIDEO CAMERA
16
VIDEO CAMERA
18
VIDEO CAMERA
DISPLAY DEVICE
21
VIDEO CAPTURE COMPONENT
20
VIDEO TRANSFER COMPONENT
22
24
COMPUTING & STORAGE DEVICE
RECORDING COMPRESSING ARCHIVING UNIT
38
VIDEO ANALYSIS UNIT
26
OBJECT TRACKING MODULE
28
VIDEO ARCHIVES
40
APPLICATION DRIVER
30
EVENTS DATABASE
42
DATABASE HANDLER
32
ALARM DISTRIBUTION UNIT
44
USER INTERFACE
34
DSA APPLICATION & SETUP PARAMETERS
46
SETUP & CONTROL DSA APPLICATION
36
DETECTION SURVELLIANCE ALARM (DSA) APPLICATION
48

FIG. 1

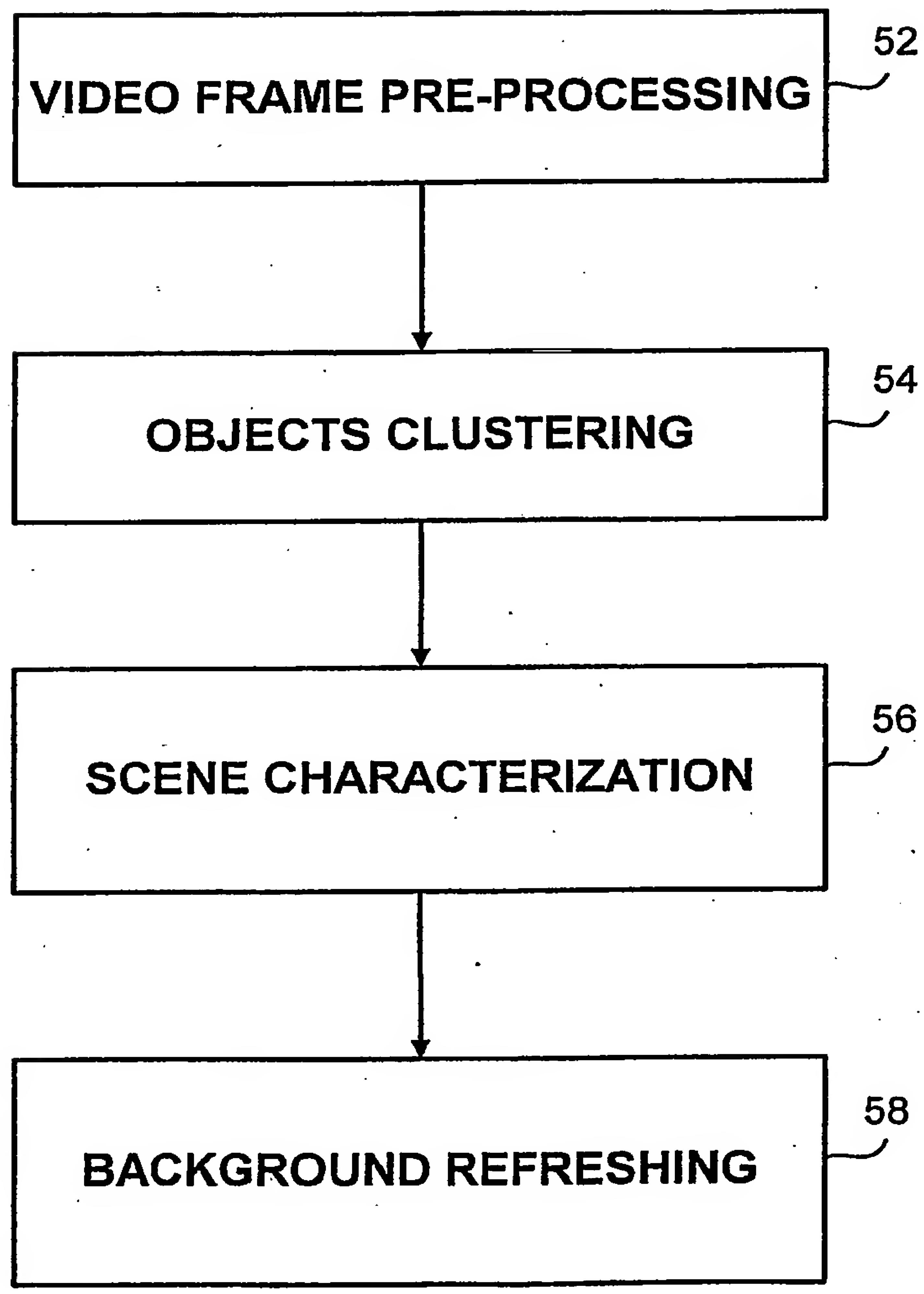

VIDEO FRAME PRE-PROCESSING
52
OBJECTS CLUSTERING
54
SCENE CHARACTERIZATION
56
BACKGROUND REFRESHING
58

FIG. 2

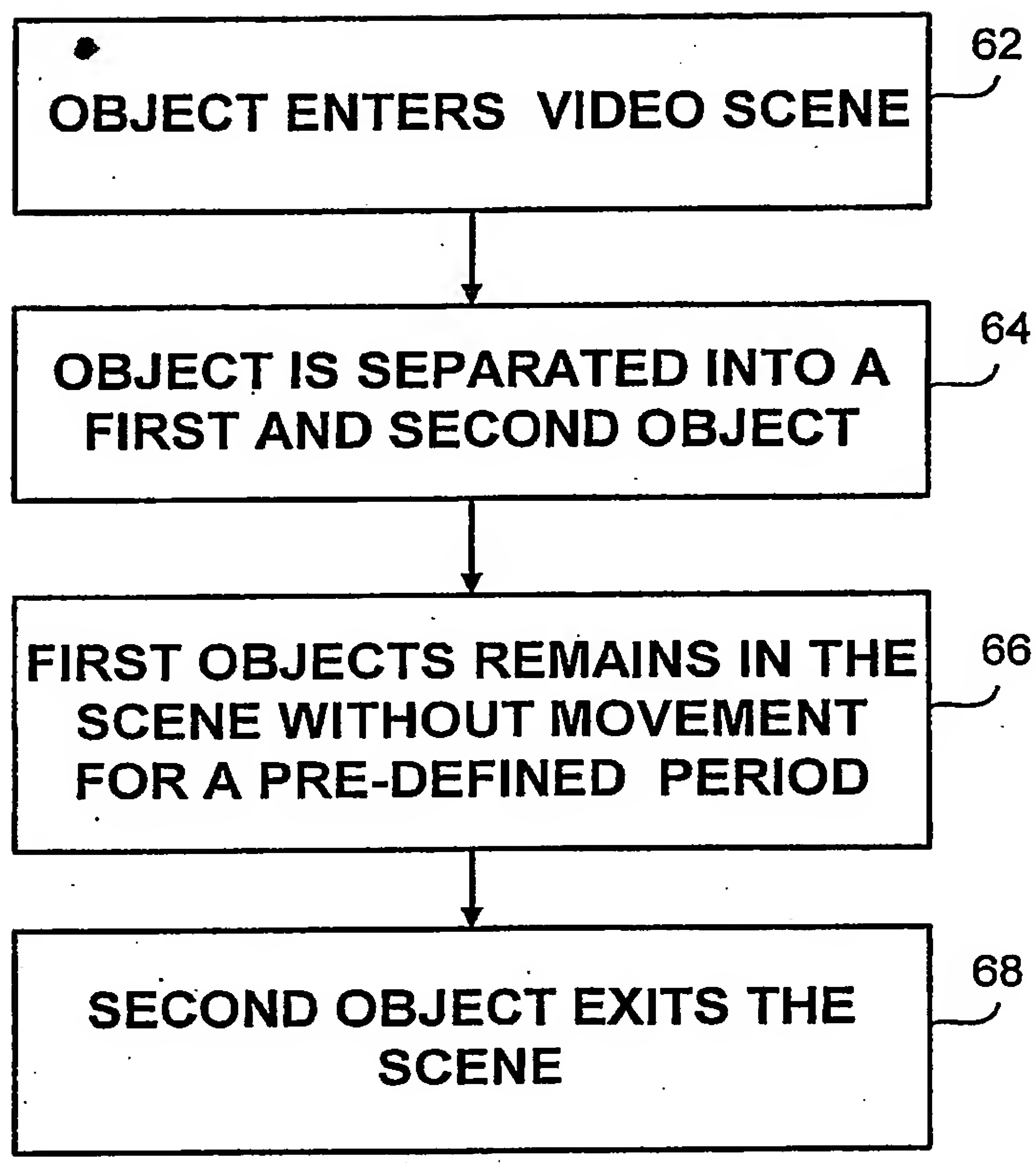

OBJECT ENTERS VIDEO SCENE
62
OBJECT IS SEPARATED INTO A FIRST AND SECOND OBJECT
64
FIRST OBJECTS REMAINS IN THE SCENE WITHOUT MOVEMENT FOR A PRE-DEFINED PERIOD
66
SECOND OBJECT EXITS THE SCENE
68

FIG. 3A

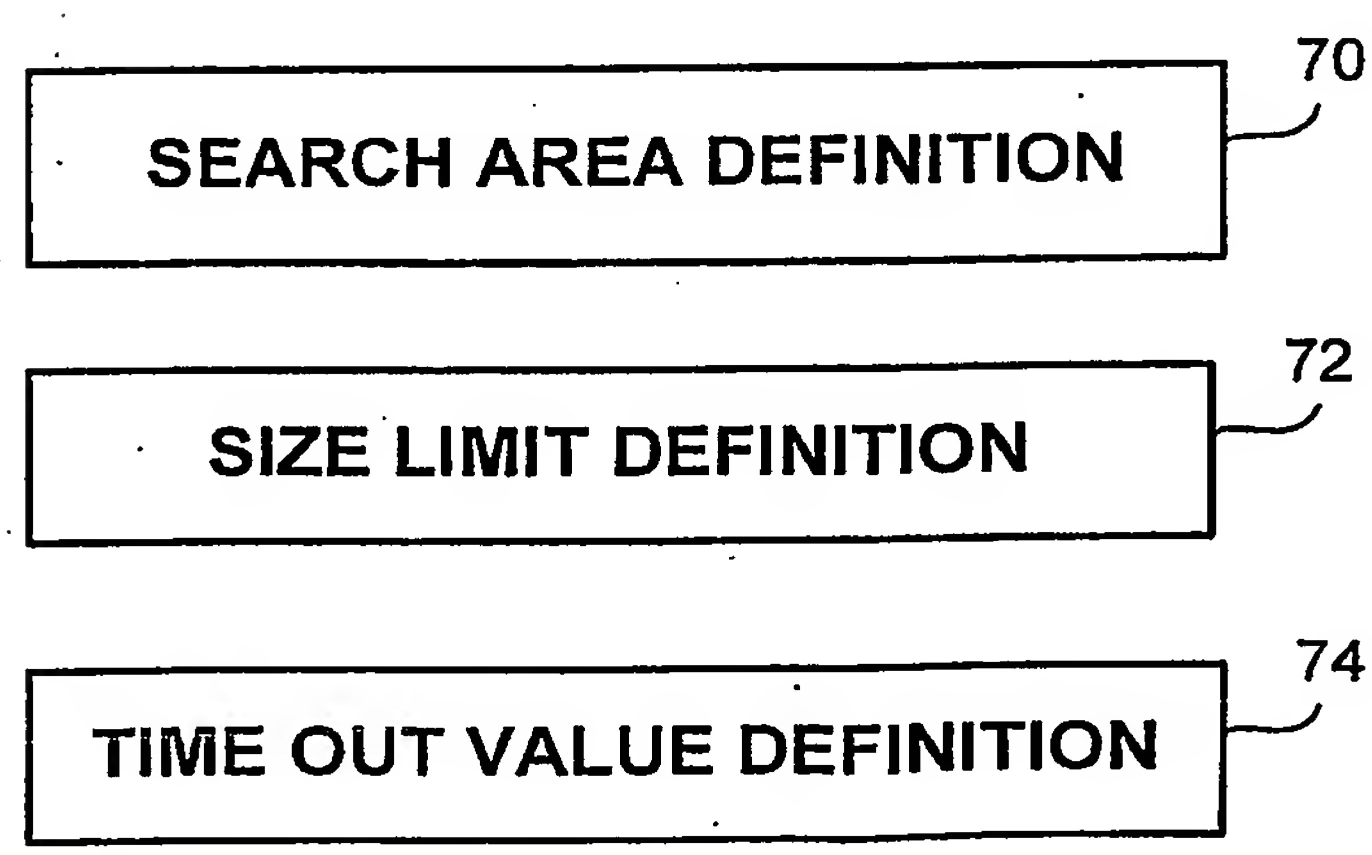

SEARCH AREA DEFINITION
70
SIZE LIMIT DEFINITION
72
TIME OUT VALUE DEFINITION
74

FIG. 3B

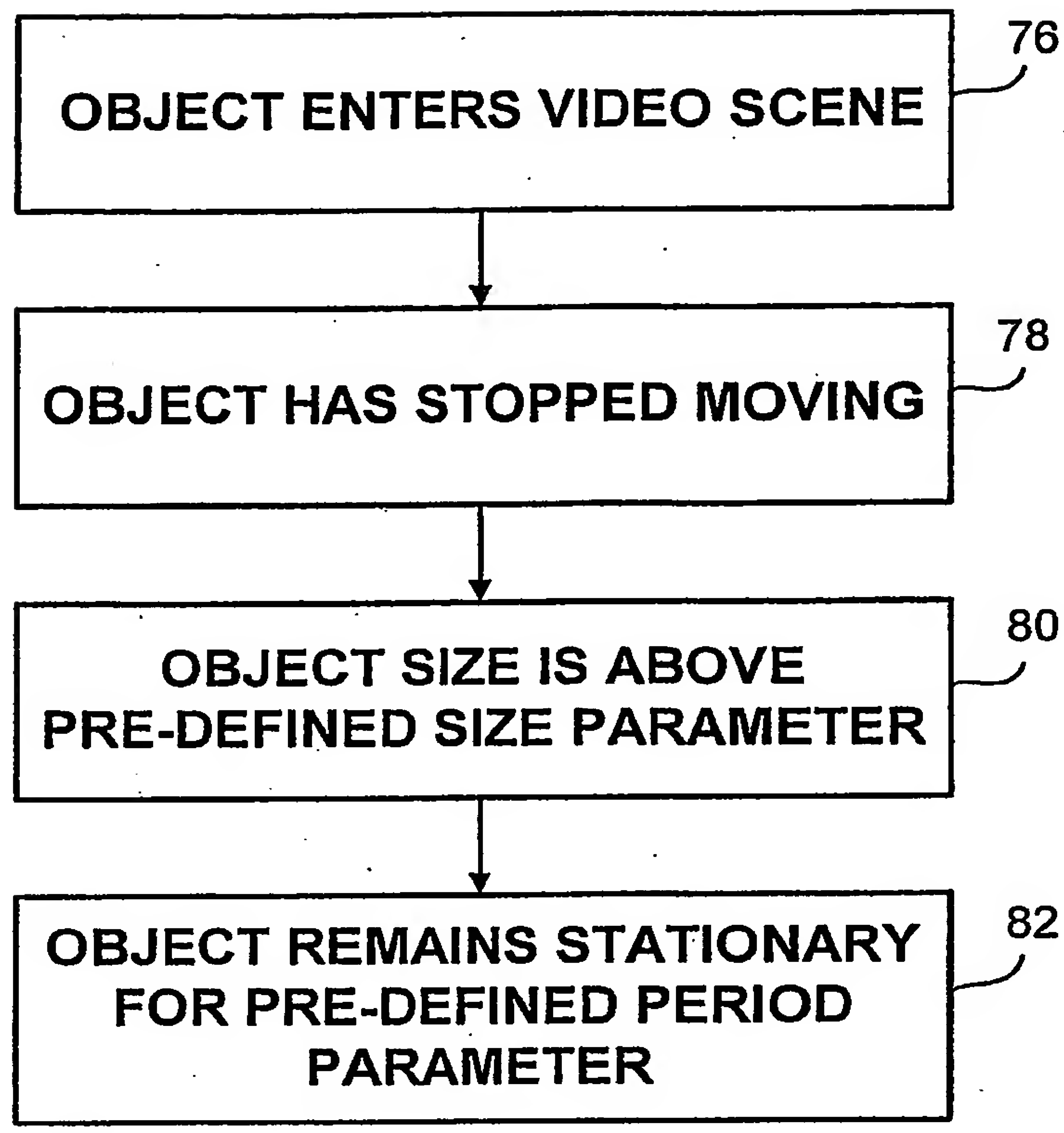

OBJECT ENTERS VIDEO SCENE
76
OBJECT HAS STOPPED MOVING
78
OBJECT SIZE IS ABOVE PRE-DEFINED SIZE PARAMETER
80
OBJECT REMAINS STATIONARY FOR PRE-DEFINED PERIOD PARAMETER
82

FIG. 4A

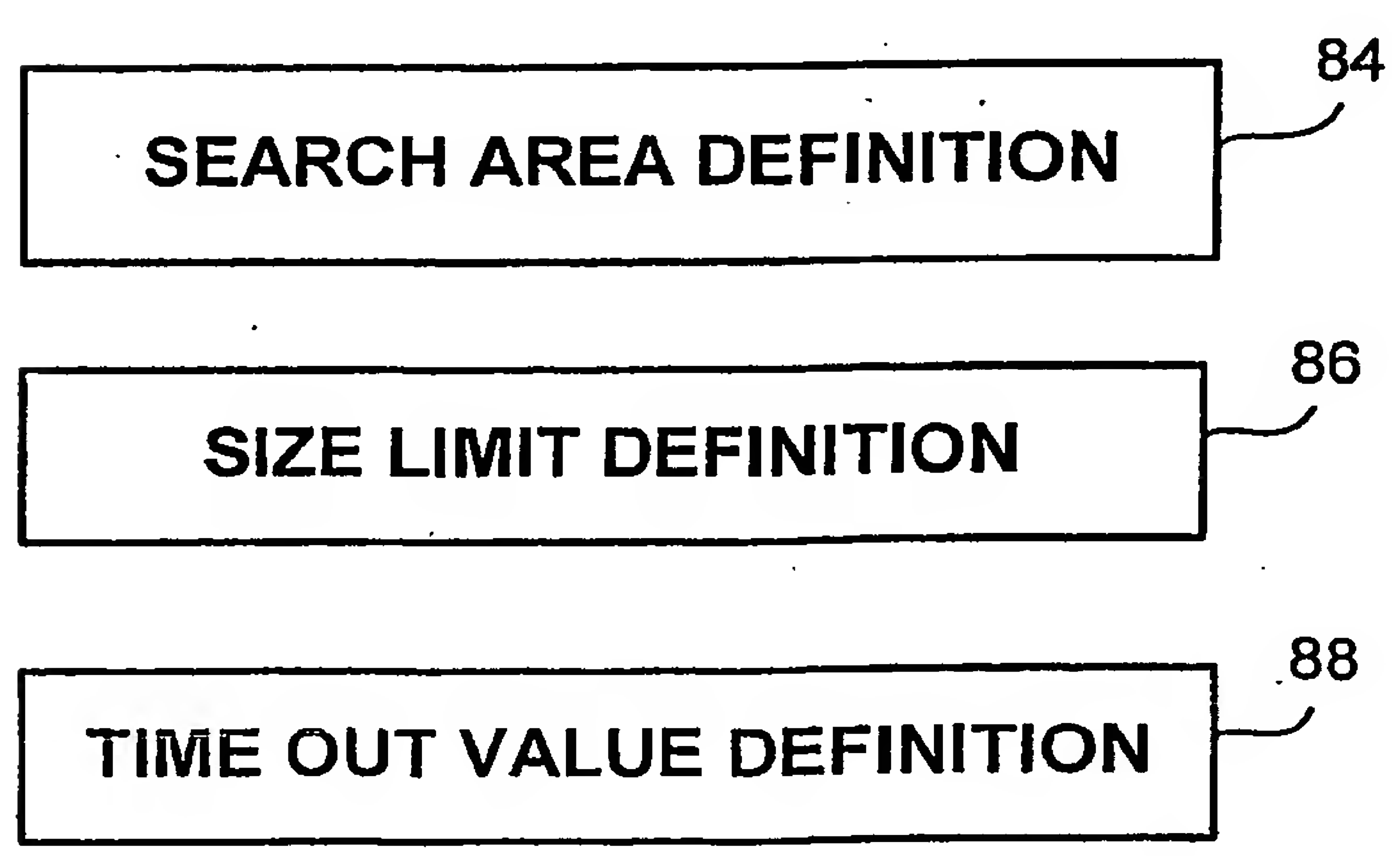
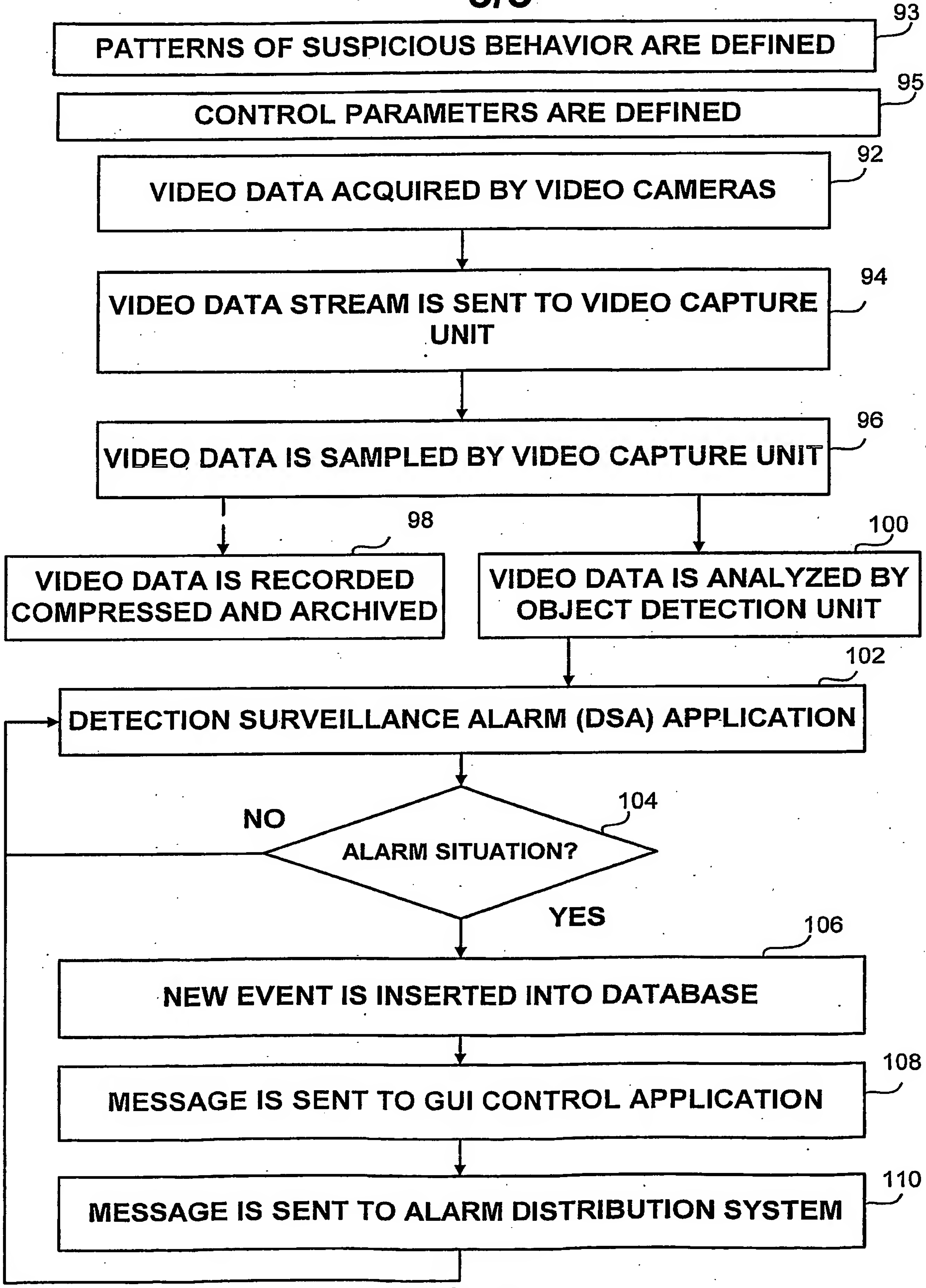

SEARCH AREA DEFINITION
84
SIZE LIMIT DEFINITION
86
TIME OUT VALUE DEFINITION
88

FIG. 4B



PATTERNS OF SUSPICIOUS BEHAVIOR ARE DEFINED
93
CONTROL PARAMETERS ARE DEFINED
95
VIDEO DATA ACQUIRED BY VIDEO CAMERAS
92
VIDEO DATA STREAM IS SENT TO VIDEO CAPTURE UNIT
94
VIDEO DATA IS SAMPLED BY VIDEO CAPTURE UNIT
96
VIDEO DATA IS RECORDED COMPRESSED AND ARCHIVED
98
VIDEO DATA IS ANALYZED BY OBJECT DETECTION UNIT
100
DETECTION SURVEILLANCE ALARM (DSA) APPLICATION
102
ALARM SITUATION?
104
NO
YES
NEW EVENT IS INSERTED INTO DATABASE
106
MESSAGE IS SENT TO GUI CONTROL APPLICATION
108
MESSAGE IS SENT TO ALARM DISTRIBUTION SYSTEM
110

FIG. 5

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G08B 15/00, G07F 7/10, G07G 3/00

A1

(11) International Publication Number: WO 98/01838

(43) International Publication Date: 15 January 1998 (15.01.98)

(21) International Application Number: PCT/US97/12000

(22) International Filing Date: 10 July 1997 (10.07.97)

(30) Priority Data:
08/677,513 10 July 1996 (10.07.96) US
08/889,889 9 July 1997 (09.07.97) US

(71)(72) Applicant and Inventor: SULLIVAN, Patrick, J. [NZ/US]; 3500 Mount Vernon Way, Plano, TX 75025 (US).

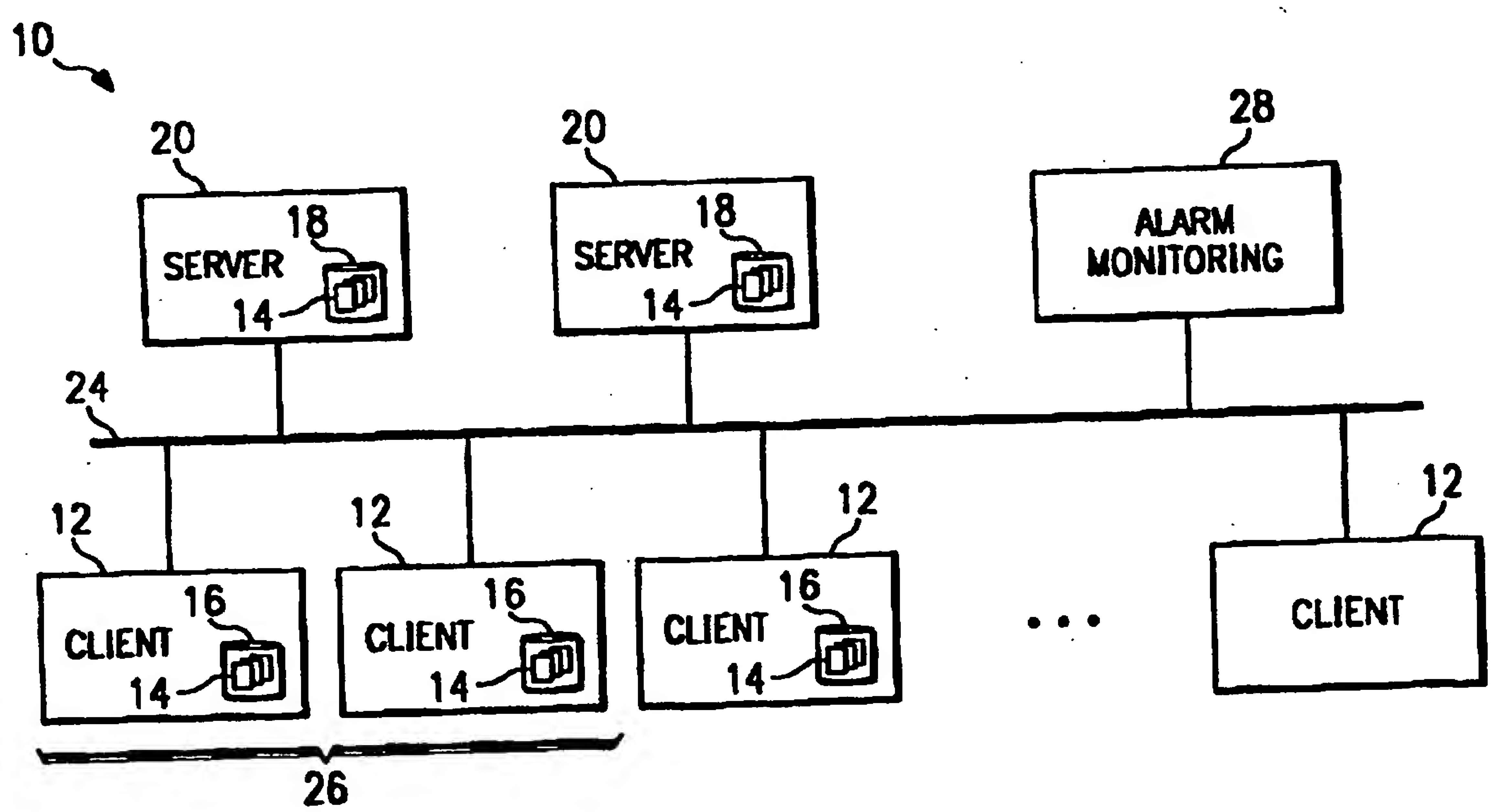
(72) Inventor: SAWAF, Ziad; 17827 Windflower Way #109, Dallas, TX 75252 (US).

(74) Agent: SHOWALTER, Barton, E.; Baker & Botts, L.L.P., 2001 Ross Avenue, Dallas, TX 75201-2980 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: VIDEO SURVEILLANCE SYSTEM AND METHOD



(57) Abstract

A video surveillance system includes clients coupled to servers using a network. The clients generate digital files that may include data, video, and audio. Using the network, the clients communicate the digital files to the servers for monitoring, analyzing, and reporting on the financial transactions occurring at the clients. The clients may also communicate in real-time the data, video, and optionnally audio to the servers. The video surveillance system also supports receiving stored data upon connection between the clients and the servers.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

VIDEO SURVEILLANCE SYSTEM AND METHOD

TECHNICAL FIELD OF THE INVENTION

This invention relates in general to surveillance and communication systems, and more specifically to a video surveillance system and method.

BACKGROUND OF THE INVENTION

A point-of-sale (POS) device, an automated teller machine (ATM), or other similar device generates data associated with a financial transaction. For example, a POS device may generate data associated with the sale of an item, whereas an ATM may generate data associated with a cash withdrawal by a customer. Due to human error, intentional misconduct, or machine malfunction, there may be a desire to display or analyze events associated with these financial transactions.

Existing surveillance systems provide some monitoring of financial transactions. For example, some surveillance systems capture data associated with financial transactions for later analysis and reporting. Other surveillance systems store video images on videotape for later visual analysis and reporting of the event. Still other systems associate or overlay financial transaction data with video stored on videotape.

SUMMARY OF THE INVENTION

In accordance with the present invention, the disadvantages and problems associated with surveillance systems have been substantially reduced or eliminated. In particular, the present invention

provides a video surveillance system and method that combines data, video, and optionally audio associated with a financial transaction in a digital file. A server retrieves the digital file from the client and provides a graphical interface to retrieve the data, video, and optionally audio from the digital file for presentation and analysis. In another embodiment, the present invention provides a video surveillance system and method that combines data, video, and optionally audio associated with a financial transaction and transmits this information in real-time to a server. A server displays the data in a data window which is overlaid on a video image corresponding to the data.

In accordance with one embodiment of the present invention, a video surveillance system includes a client that generates data associated with a financial transaction. The client has a camera that generates video associated with the financial transaction. The client stores data and video in a digital file. A server is coupled to the client using a communications network and receives the digital file from the client and stores the digital file in a memory. The server has a graphical interface that retrieves data and video from the digital file stored in the memory for presentation.

In accordance with another embodiment of the present invention, a video surveillance system includes a client that generates data associated with a financial transaction. The client has a camera that generates video associated with the financial transaction. The client transmits the data and video over a communication network. A server is coupled to the client using the communications network and receives the data and video from the client. The server displays the video and data in real-time.

Important technical advantages of the present invention include the storage and/or real-time transmission and viewing of data, video, and optionally audio associated with a financial transaction. In particular, a client, such as a point-of-sale (POS) device like a cash register or an automated teller machine (ATM), generates data associated with and/or upon the occurrence of a financial transaction. The client also includes a camera that generates associated video and optionally a microphone that generates associated audio. The client may store data, video, and audio in a single multimedia digital file. In a particular embodiment, the client includes two modes of operation. In the first mode, the client includes only data associated with the financial transaction in the digital file. In the second mode associated with an exception condition of the financial transaction, the client includes data, video, and optionally audio in the digital file. The exception condition may be defined by information transmitted from the server.

The storage of different information into a digital file provides several important technical advantages. The digital file may be formatted, compressed, and communicated using digital communications technology. The digital file may be scrambled, rearranged, encoded, or otherwise processed to prevent tampering or disassociation of data, video, and audio. Also, a digital file format allows more sophisticated database storage, retrieval, and reporting functions.

Another important technical advantage of the present invention includes a server coupled to the client that retrieves the digital file. The server includes a graphical interface that allows a user of the server to display, analyze, and generate reports

on information contained in the digital file. In a particular embodiment, the server is coupled to a plurality of clients, and each client generates digital files for financial transactions occurring at the client. The server collects and stores these digital files in a database. The graphical interface accesses the database to allow selection, presentation, analysis, and reporting of the financial transactions represented by the digital files. In a particular embodiment, data for each financial transaction appears as an entry in a table of financial transactions. Highlighted entries may indicate the existence of video associated with the data. The graphical interface may also include a video window for viewing associated video and a search/report window to allow selection and analysis of financial transactions.

In another embodiment, the client provides a further technical advantage by transmitting data, video, and audio across a communications network in real-time. Additionally, data can be transmitted from the client to the server upon initialization of a real-time connection. This data may represent daily sales total, transaction totals, number of items sold since last contact or some other information. In this embodiment, a server coupled to the client receives the transmitted data and video. The server includes a display that allows the data to be shown as a data window overlaid on the associated video. Multiple data windows as well as multiple video windows can be displayed. In a particular embodiment, the server is coupled to a plurality of clients and each client transmits data and video for financial transactions occurring at the client. The server displays these transactions in multiple windows. An operator can change the views in the windows or the windows can be

changed automatically, based on some preexisting criteria. Other technical advantages are readily apparent to one skilled in the art from the following figures, descriptions, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention, and for further features and advantages, reference is now made to the following written description taken in conjunction with the accompanying drawings, in which:

FIGURE 1 illustrates a video surveillance system;

FIGURE 2 illustrates a client in the video surveillance system;

FIGURE 3 illustrates a server in the video surveillance system;

FIGURE 4 illustrates a graphical interface at the server in the video surveillance system;

FIGURE 5 illustrates the components of an exemplary digital file used in the video surveillance system;

FIGURE 6 is a flowchart of a method of operation of the client in the video surveillance system;

FIGURE 7 is a flowchart of a method of operation of the server in the video surveillance system;

FIGURE 8 illustrates a video surveillance system capable of real-time transmission of video and data;

FIGURE 9 illustrates a display which includes a video window and one or more data windows;

FIGURE 10 illustrates a display divided into multiple video windows and data windows;

FIGURE 11 is a flowchart of real-time data and video transmission; and

FIGURE 12 is a flowchart for a method of updating data from a client.

DETAILED DESCRIPTION OF THE INVENTION

FIGURE 1 illustrates a video surveillance system 10 that includes clients 12 coupled to servers 20 using a communications network 24. In operation, clients 12 generate digital files 14 that include data, video, and optionally audio associated with financial transactions. Clients 12 communicate digital files 14 to servers 20 using network 24. Servers 20 store digital files 14 received from clients 12 in databases 18, and provide remote monitoring, reporting, and analysis of financial transactions occurring at clients 12.

Clients 12 may include or be associated with any electronic device that generates data on a financial transaction, such as a point-of-sale (POS) device like a cash register, automated teller machine (ATM), or any other appropriate device that generates data on a financial transaction. Clients 12 may be located at one or more sites, associated with one or more business organizations, or otherwise arranged or grouped in any appropriate manner. For example, two or more clients 12 may be co-located at a site, operated by the same business organization, or otherwise associated as indicated by bracket 26. Each server 20 in system 10 receives digital files 14 associated with financial transactions occurring at one or more designated or associated clients 12. System 10 contemplates any association or arrangement of clients 12 and servers 20 to accomplish remote monitoring and analysis of financial transactions.

Network 24 represents hardware and software used in any suitable communications network or computer network, such as a local area network (LAN), wide area network (WAN), public switched telephone network, integrated services digital network (ISDN), switched-56 telephone network, private branch exchange (PBX),

the global computer network known as the Internet, or any other appropriate technology or technique that allows components of system 10 to communicate information. Although client 12 and server 20 are referred to in the nomenclature of a client/server environment, it should be understood that client 12 and server 20 may be any type of computer operating in any suitable environment that communicates using network 24. Each component in system 10 includes any suitable hardware and software components to interface with and communicate using network 24.

In a particular embodiment, network 24 supports one-way and two-way audio/video conferencing. Throughout this description, audio/video conferencing includes conferencing of audio alone, video alone, or both audio and video, together with any associated data. For example, network 24 may include components to implement an integrated services digital network (ISDN) communications facility that supports the ITU H.320 video conferencing standard. In this embodiment, each component of system 10 may include appropriate transceivers, coders/decoders (codecs), interface cards, and other hardware and software to implement audio/video conferencing and underlying data transfer.

An alarm monitoring station 28 is also coupled to network 24 and detects alarm conditions at clients 12. In response to this detection, station 28 establishes communication with the particular client 12 that generates the alarm condition. Station 28 may display in a direct, dedicated, real-time, or near real-time fashion data, video, and audio generated at the particular client 12 that generated the alarm condition. Station 28 may also perform one-way or two-way audio/video conferencing with the particular client 12. In a particular embodiment, station 28

alerts and dispatches police, fire, security, or other officials to client 12.

In operation, clients 12 perform financial transactions and generate digital files 14 associated with the financial transactions for storage in databases 16. At appropriate times, server 20 receives digital files 14 from clients 12, and stores these digital files 14 in database 18. In one embodiment, database 18 maintained at server 20 includes digital files 14 collected from numerous clients 12. Server 20 includes a database management system and a graphical interface to display, select, analyze, and report on financial transactions occurring at clients 12 that correspond to digital files 14 maintained in database 18.

FIGURE 2 illustrates client 12 in more detail. On-site input/output devices 50 include microphone 52, speaker 54, cameras 56, and display 58. A video switch 60, coupled to cameras 56 and display 58, selects video from one or more cameras 56. A video cassette recorder (VCR) 62 or other appropriate recording device is coupled to input/output devices 50, and records video and audio information on videotape 64.

Input/output devices 50 are coupled to a converter 70, which passes video 72 and audio 74 in digital format to a controller 76. Controller 76 is coupled to and receives data 82 regarding a financial transaction from ATM 78, POS 80, or any other device that generates data 82 regarding a financial transaction. An alarm 83 is also coupled to controller 76, and represents a motion detector, clock, panic button, or other device that generates an alarm condition 85 at client 12.

Controller 76 is coupled to database 16 which stores digital files 14 and exception condition 84.

Exception condition 84 comprises information that directs client 12 when to store video and optional audio for particular financial transactions. For example, exception condition 84 may represent one or more activities, such as keystrokes at ATM 78 or POS 80, that when detected in data 82 triggers the capture of video 72 and/or audio 74 for the financial transaction. Exception condition 84 may be defined as a noise threshold in audio 74 or a pixel or picture variance or difference threshold in video 72 that, when exceeded, triggers the capture of video 72 and audio 74. Controller 76 is also coupled to codec 86, which in turn is coupled to network 24 using interface 88.

Particular components of client 12 may operate on one or more computers, shown generally as computer 90. Computer 90 maintains and executes the instructions to implement converter 70, controller 76, codec 86, and interface 88, and includes any suitable combination of hardware and software to provide the described function or operation of these components. Database 16 comprises one or more files, lists, or other arrangement of information stored in one or more components of random access memory (RAM), read only memory (ROM), magnetic computer disk, CD-ROM, other magnetic or optical storage media, or any other volatile or nonvolatile memory.

Computer 90 includes an input device 92 such as a keypad, touch screen, mouse, or other device, that can accept information. Output device 94, such as a computer display or speaker, conveys information associated with the operation of client 12, including digital data, visual information, or audio information. Both input device 92 and output device 94 may include fixed or removable storage media such as a magnetic computer disk, CD-ROM, or other suitable

media to both receive output from and provide input to client 12. Processor 96 and its associated memory execute instructions and manipulate information in accordance with the operation of client 12.

In operation, input/output devices 50 may operate in an analog or mixed analog/digital environment. For example, cameras 56 may generate and display 58 may display video in a standard television format such as NTSC or other analog signal format. Similarly, microphone 52 may generate and speaker 54 may convey audio information in analog form. If appropriate, converter 70 converts analog signals used by one or more input/output devices 50 into digital signals for video 72 and audio 74 used by controller 76. In one embodiment, input/output devices 50 generate and receive digital data and the operation of converter 70 is unnecessary.

Upon the occurrence of a financial transaction, ATM 78 or POS 80 generates data 82 associated with the financial transaction. Controller 76 analyzes video 72, audio 74, and/or data 82 to determine if it indicates, corresponds to, or is associated with exception condition 84 stored in database 16. In a first mode, controller 76 determines that video 72, audio 74, and/or data 82 are not associated with exception condition 84 and stores only data 82 generated by ATM 78 or POS 80 as digital file 14 in database 16. In a second mode, controller 76 determines that video 72, audio 74, and/or data 82 generated by ATM 78 or POS 80 is associated with exception condition 84, which triggers the capture of video 72 and optionally audio 74. Therefore, in the second mode of operation, controller 76 includes data, video 72, and optionally audio 74 associated with the financial transaction in digital file 14 stored in database 16.

Contemporaneously with the storage of digital file 14 in database 16 or at an appropriate later time, controller 76 retrieves one or more digital files 14 from database 16 for transmission to server 20 using codec 86, interface 88, and network 24. Client 12 may schedule delivery of digital files 14 to server 20 in any appropriate manner. For example, client 12 may communicate digital files 14 to server 20 at off-peak hours, at the end of a shift, at specified intervals during a day, week, or month, or at any other appropriate time, depending on the particular requirements of the business organization operating client 12. In addition, client 12 may initiate communication of digital files 14 in response to a command received from server 20 over network 24. Also, alarm condition 85 generated by alarm 83 may cause client 12 to immediately communicate digital file 14 associated with alarm condition 85. In this embodiment, client 12 may transmit alarm condition 85 to station 28 and establish a direct, dedicated, real-time, or near real-time one-way or two-way audio/video conference with station 28.

In combination with or separate from the generation and communication of digital files 14, client 12 also supports one-way and two-way audio/video conferencing using network 24. For one-way audio/video conferencing, converter 70 passes video 72 from cameras 56 and audio 74 from microphone 52 to controller 76. Controller 76 and codec 86 place video 72 and audio 74 into an appropriate format such as the video conferencing standard described in ITU H.320. Controller 76 may also include any data generated at client 12 in the conferencing information. In a particular embodiment, one-way audio/video conferencing signals are multiplexed and compressed onto a single digital bit data stream and

transmitted to server 20 or station 28 using the ISDN communications standard supported by network 24.

For two-way audio/video conferencing, the components of client 12 perform the same outgoing conferencing capability, but also receive audio/video conferencing signals from server 20 using network 24, interface 88, and codec 86. Controller 76 receives incoming signals from codec 86, separates the signals, and passes video 72 and audio 74 to converter 70. Converter 70 performs conversion, if appropriate, and presents incoming conferencing signals to speaker 54 and display 58. Controller 76 may also extract data from the incoming conferencing signals.

In a particular embodiment, real-time video 72 and optionally audio 74 is sent along with corresponding data 82 generated by ATM 78 or POS 80. The term "real-time" means real-time, near real-time, or contemporaneous as possible but subject to limitations in communication systems that cause substantial time to elapse between the capturing of video 72 and data 82 and the display at server 20. In this embodiment, video 72 and optionally audio 74 is sent to controller 76 where it combines with data 82 from ATM 78, POS 80 or any other device that generates data 82 regarding a financial transaction. Alternatively, instead of controller 76, any other device can be used that can combine video 72 and data 82. Data 82 from more than one ATM 78 or POS 80 can be transmitted. Additionally, multiple video windows can be transmitted, each one representing a different camera 56 feed. These types of transmissions can also occur at multiple clients 12. Video 72 is transmitted along with the corresponding data 82 over network 24 via interface 88. Alternatively, data 82 can be stored in database 16 over a period of time. Upon establishment of network connection or some other

occurrence, data 82 is transferred from client 12 to server 20. Server 20 can then query either its own database 18 or database 16 at server 12.

FIGURE 3 illustrates server 20 in more detail. Input/output devices 100 include camera 102, microphone 104, and speaker 106. Input/output devices 100 are coupled to a codec 108, which in turn is coupled to network 24 using an interface 109. A controller 110 is coupled to codec 108, display 112, and input devices 113. Display 112 displays information contained in digital files 14 received from clients 12. In particular, display 112 presents a graphical interface 116 that allows a user of server 20 to display, select, analyze, and report on financial transactions occurring at clients 12 that correspond to digital files 14 maintained in database 18. Also included in database 18 is a data configuration 15 which allows data to be overlaid on display 112 in a variety of formats. Input devices 113 may include a keyboard, mouse, other pointing device, or any other appropriate input device that allows the user to interact with graphical interface 116 and direct the operation of server 20.

Controller 110 is also coupled to database 18, which stores digital files 14 received from clients 12. Alternatively, a video cassette recorder 107 can be used to store real-time video 72. Database 18 includes a database management system 114 that provides traditional database features to store, retrieve, and manipulate information stored as digital files 14 for monitoring, analyzing, and reporting on financial transactions occurring at clients 12. Database management system 114 supports any suitable flat file, hierarchical, relational, object-oriented, or parallel database operation.

In operation, server 20 receives data 82, video 72, and optionally audio 74 in the form of digital files 14 from network 24 using interface 109. Codec 108 decompresses and converts this information into a proper format for storage in database 18 by controller 110. The retrieval of digital files 14 from clients 12 may occur on a periodic basis defined by clients 12, a periodic basis defined by server 20, or as a result of server 20 polling clients 12 with commands to download information.

In response to a request from controller 110, database management system 114 accesses selected digital files 14 and passes this information to controller 110 for presentation by graphical interface 116 on display 112. Using graphical interface 116, the user can display, select, view, analyze, and report on information associated with financial transactions occurring at clients 12.

In an alternative embodiment, server 20 receives data 82 and video 72 from network 24 in real-time. Video 72 is displayed on display 112 based on data configuration 15 stored in database 18. Overlaid as a data window on display 112 is a representation of data 82, such as a cash register receipt. Alternatively, multiple data windows can be displayed corresponding to data 82 from different ATM 78 or POS 80 at the same or different location. Multiple video windows, each one a different video 72 from a different camera 56 at the same or different location can be shown on display 112. Data windows can be displayed for each video 72. A user can switch video 72 based on what is occurring in a data window or views can be switched automatically based on some preexisting criteria. FIGURES 9-11 describe the techniques to display video 72 and data 82 in a variety of arrangements.

Additionally, upon connection between client 12 and server 20, client 12 can automatically transfer data 82 to server 20. This can be all the financial records since last connection, all the financial records over a certain period of time, or some other configuration. Server 20 can then query either its own database 18 or the client's database 16 for further information. Data configuration 15 can control the display of data 82. For example, data configuration 15 can have the display show total sales for a certain time period broken down by categories of items purchased. FIGURE 12 describes in more detail techniques for updating data from a client.

FIGURE 4 illustrates in more detail the components of graphical interface 116. Graphical interface 116 includes a table 120 having a number of entries 122 associated with financial transactions. Each entry 122 includes all or a portion of data 82 generated by POS 78 or ATM 80 at client 12. Highlighted entries 124 may be emphasized by shading, font changes, color differences, or other appropriate technique to indicate the existence of associated video and/or audio. In a particular embodiment, entries 122 correspond to data received from clients 12 operating in a first mode in which digital file 14 includes data 82, and highlighted entries 124 correspond to information retrieved from clients 12 operating in a second mode in which digital file 14 includes data 82, video 72, and optionally audio 74. In this manner, table 120 provides data 82 on associated financial transactions, and also conveys visually those financial transactions associated with particular defined exception conditions 84 at clients 12. Highlighted entries 124 may then be quickly recognized by the user of server 20 and analyzed as a

suspect or more closely monitored financial transaction.

Upon a user selecting highlighted entry 124 using input device 113, video window 126 presents a still frame or a selected portion of partial or full motion video 72 associated with the selected highlighted entry 124. Optional audio 74 may also be presented simultaneously with video 72. The user may manipulate a toolbar 128 to play, pause, stop, fast forward, rewind, adjust the volume, or perform other appropriate functions to analyze video 72 and audio 74 presented by graphical interface 116. Furthermore, video window 126 may present a magnification box 130 that allows the user to analyze selected portions of video 72 in more detail using zoom, pan, and other functions. The storage of vide 72 and audio 74 as digital information enables more sophisticated analysis techniques, such as the techniques provided by toolbar 128 and magnification box 130.

Graphical interface 116 also includes a search/report window 132 that allows the user of server 20 to specify particular financial transactions to view in table 120. For example, search/report window 132 may prompt the user for a number of parameters that specify the desired financial transactions to view. Parameters may include time, date, store identifier, register identifier, amount of transaction, all transactions involving a particular item, all transactions meeting an exception condition, or any other appropriate parameter. Search/report window 132 may also include various printing, reporting, and analyzing capabilities of server 20.

FIGURE 5 illustrates digital file 14 generated by client 12, optionally stored in database 16 at client 12, and stored in database 18 at server 20. Digital file 14 includes data 82 generated by ATM 78 or POS 80

at client 12, video 72, and audio 74. As described above, client 12 operating in a first mode in which exception condition 84 is not met may only include data 82 in digital file 14. However, in a second mode in which exception condition 84 is met, client 12 may include data 82, video 72, and optionally audio 74 in digital file 14.

Data 82 includes a transaction identifier 200, date and time 202, POS or ATM identifier 204, and site identifier 206, which may all be considered together as identifiers 208 that uniquely specify digital file 14 in system 10. Data 82 also includes transaction data 210, that may specify transaction type, item identification, item cost, taxable amount, amount tendered, tax added, total, withdrawal amount, account information, user information, keys depressed at ATM 78 or POS 80 during the financial transaction, or any other data associated with the financial transaction. Controller 76 at client 12 may analyze transaction data 210 to determine if exception condition 84 is met. Data 82 may also include other data 212, such as a measure of the time the cash register door is open, an identifier of the employee on duty, an estimate of the number of persons in the store, or other information not directly related to the financial transaction but provided in data 82 for further analysis of the financial transaction. For clarity, digital file 14 in FIGURE 5 arranges information in blocks. However, data 82, video 72, and audio 74 may be arranged in any format or order, depending upon the particular implementation and technology used in system 10.

The maintenance of data 82, video 72, and audio 74 in a single digital file 14 provides several technical advantages. Digital file 14 may be scrambled, rearranged, encoded, or otherwise processed

to prevent tampering or disassociation of data 82 with its corresponding video 72 and audio 74. In addition, system 10 capitalizes on digital storage, compression, and communications techniques to quickly and efficiently gather digital information at server 20. Also, the digital format of digital file 14 enables more sophisticated database storage, retrieval, and reporting functions to be performed at server 20.

FIGURE 6 illustrates a flow chart of a method of operation of client 12. The method begins at step 300 where the next financial transaction occurs at ATM 78 or POS 80. ATM 78 or POS 80 generate data 82 associated with and upon the occurrence of the financial transaction at step 302. Continuously, upon the occurrence of the financial transaction at step 300, or at any appropriate time, microphone 52 and cameras 56 generate audio/video information associated with financial transaction at step 304. Throughout this description, the term audio/video refers to video alone, audio alone, or both video and audio. If appropriate, converter 70 may generate video 72 and audio 74 for communication to controller 76.

If controller 76 determines that exception condition 84 stored in database 16 is not met at step 306, client 12 enters a first mode and stores digital file 14 including data 82 in database 16 at step 308. If exception condition 84 is met at step 306, client 12 enters a second mode and stores digital file 14 including data 82, video 72, and optionally audio 74 in database 16 at step 310.

If controller 76 detects alarm condition 85 from alarm 83 at step 312, client 12 establishes communication with server 20 or optionally alarm monitoring station 28 at step 314. While client 12 maintains alarm condition 85, client 12 and server 20 or station 28 exchange data, video, and audio at step

316 to implement a one-way or two-way audio/video conferencing link for remote surveillance, management, or supervision. If alarm condition 85 persists at step 318, client 12 and server 20 or station 28 continue to exchange data 82, video 72, and audio 74 at step 316. If alarm condition 85 is over at step 318 and the operation of client 12 is not done at step 320, the method returns to process the next financial transaction at step 300.

If no alarm condition 85 exists at step 312, client 12 determines whether to send digital files 14 to its associated server 20 at step 322. If digital files 14 are to be sent to server 30, client 12 selects digital files 14 to send at step 324. This may be done in response to a command received from server 20 or by specifying locally by client 12 or remotely by server 20 various parameters, such as time, site identifier, register identifier, or other appropriate parameter to select digital files 14. Controller 76 passes selected digital files 14 to codec 86 for formatting at step 326. Interface 88 using network 24 sends digital files 14 to server 20 at step 328. If appropriate, client 12 updates database 16 at step 330, for example, by deleting digital files 14 transmitted to server 20. If the operation of client 12 is not done at 320 or client 12 determines not to send digital files 14 to server 20 at step 322, the method returns to process the next financial transaction at step 300.

FIGURE 7 illustrates a flow chart of a method of operation of server 20. The method begins at step 400 where server 20 receives digital files 14 from associated clients 12. Server 20 stores the received digital files 14 in database 18 at step 402. Dashed feedback arrow 404 indicates that steps 400 and 402 may execute in parallel, in series, in the background,

or in any other appropriate manner to receive digital files 14 from clients 12.

A user of server 20 may input information into search/report window 132 of graphical interface 116 to select digital files 14 stored in database 18 to view at step 406. In response to the user's request, database management system 114 retrieves selected digital files 14 from database 18 and passes this information to controller 110 at step 408. Graphical interface 116 presents data 82 associated with retrieved digital files 14 as entries 122 and 124 in table 120 at step 410. Graphical interface 116 also highlights particular entries 124 with associated video 72 and optional audio 74 at step 412.

Upon selecting particular digital files 14 from database 18 and presenting entries 122 and 124 in table 120, graphical interface 116 of server 20 supports several user functions as illustrated by branch 414. Graphical interface 116 may support these functions in parallel, in serial, or in any other fashion to allow interaction with the user of server 20. If the user selects highlighted entry 124 at step 416, graphical interface 116 presents associated video 72 and audio 74 in video window 126 at step 418. Graphical interface 116 then services functions of toolbox 128 and magnification box 130 at step 420 to allow further analysis of video 72 and audio 74 associated with the selected highlighted entry 124.

If the user selects an analysis or reporting function at step 422, graphical interface 116 services the analysis and reporting function at step 424. For example, the user of server 20 may request summary statistics, print information, run predefined reports, or perform any other function on data 82 displayed in table 120 or maintained as digital files 14 in database 18. Graphical interface 116 outputs the

results of the analysis and reporting functions at step 426.

The user may also request a new table at step 428 by submitting another query in search/report window 132. In response, the method continues at step 406 where server 20 selects digital files 14 to view. If the operation of server 20 is not done at step 430, server 20 continues to receive digital files 14 at step 400 and store digital files 14 in database 18 at step 402.

FIGURE 8 illustrates a video surveillance system 10 capable of real-time transmission of video 72 and data 82. Clients 12 are coupled to servers 20 via a communication network 24 as previously discussed in conjunction with FIGURE 1. Each server 20 includes display 112 which is capable of displaying real-time video 72 and data 82 received from client 12. In this embodiment, video 72 is sent by client 12 over connection 24 as soon as it is captured by camera 56. This is called real-time video, although it is understood that the limits of communication network 24 and other components of surveillance system 10 may introduce an appreciable delay between the capturing of video 72 and its display at server 20. Data 82 corresponding to video 72 is also sent along network 24. Additional data 82 from other sources at that location can be sent along network 24 as can additional video 72 from other cameras.

FIGURE 9 illustrates display 112 which includes video window 126 and one or more data windows 127. Data window 127 represents an overlay of data 82 from ATM 78 or POS 80 or any other device capable of producing data 82 generated by a financial transaction. For example, data window 127 may represent a live or scrolling version of a cash register receipt corresponding to the image of a

customer purchasing goods in a store in video windows 126. Multiple data windows 127 can be overlaid on a video window 126, each corresponding to an ATM 78 or POS 80 at the same or different location. An operator at server 20 can switch video 72 corresponding to a particular data window 127 or the switching can be done periodically or automatically, based on a variety of criteria such as an alarm condition or preprogramed response. These include switching to video 72 based on data 82 (e.g., the presence of data 82, amount of money spent, key depressions at ATM 78 or POS 80, etc.), type of item purchased, movement in a video window, sound levels, or any other suitable criteria. Server 20 may switch, arrange, or display video window 126 and/or data window 127 in response to these criteria.

FIGURE 10 illustrates a display 112 divided into multiple video windows 126. These video windows 126 can be from cameras 56 at the same or different locations. Data windows 127 are overlaid in each video window 126. Multiple data windows 127 can be displayed in a given video window 126 as discussed in FIGURE 9. Multiple video windows 126 and data windows 127 may be repositioned, resized, or otherwise manipulated or arranged in response to criteria discussed above with reference to FIGURE 9.

FIGURE 11 is a flowchart of real-time video 72 and data 82 transmission. In step 500, client 12 generates data 82 from a financial transaction at ATM 78 or POS 80 or other device. Client 12 also generates video 72 corresponding to data 82 at step 502. Client 12 transmits video 72 and data 82 over network 24 to server 20 in step 504. Video 72 and data 82 are communicated in real-time and can be from multiple sources. In step 506, data is overlaid as a data window 127 with video 72 on display 112. Server

20 overlays a single or multiple data windows 127 on a single or multiple video window 126 in any appropriate configuration as discussed above with reference to FIGURE 9 and 10. Step 508 determines if a change to a display needs to be made based on a variety of criteria. Manual changes, such as those initiated by an operator, are covered in step 510. These would include an operator switching to a window based on a transaction appearing in a data window or an operator switching camera views as part of a normal scan pattern. In step 512, server 20 automatically configures windows based on changes in video 72. For example, sudden movement may trigger a video window to appear. Also, if a camera becomes obstructed, a change in a video window might be triggered. In step 514, display 112 automatically changes due to the presence or content of data 82, such as the amount of purchase, the type of purchase, some alarm condition, keystrokes, or other criteria.

FIGURE 12 is a flowchart for a method of updating data 82 from client 12. In step 520, a link between client 12 and server 20 is established. This step can involve the actual establishment of a link over network 24, the restoring of a paused link, initializing the update procedure over an already established link, or some other connection criteria or prearranged transfer time. In step 522, client 12 communicates data to server 20. This data 82 can contain register or item totals for a given time period, a running total since last connection, the total number of certain items sold, raw inventory or transaction data, or any other numeric or alphanumeric data that conveys information on the activity at client 12. Data 82 is typically stored in a file at client 12 and information gathered over a period of

time is accumulated in that file. This file can be stored at server 20 upon receipt.

In step 524 server 20 displays data 82 in a format determined by data configuration file 15 stored in database 18. The display format may specify grand totals, sales on items of interest, inventory, cash on hand, or some other information. Configuration file 15 may be stored as an initialization file or some type of configuration file. Configuration file 15 is designed to be updated easily to allow displays to be designed efficiently. Server 20 maintains multiple configuration files 15 based on client 12, the identity of the operator at client 12, the identity of the operator at server 20, a time measure (e.g., time of day, day of the week, day of the month, quarter, etc.), a particular report format, or any other appropriate criteria.

In step 526, the process determines if additional queries need to be made. Additional queries can be made to extract data not transferred or to display data in an alternative format. If so, step 528 determines if the queries are to be made locally, that is, at the server. If so, server 20 performs queries on the server's database 18 in step 530. If not, server 20 performs queries on the client's database 16 in step 532.

Although the present invention has been described in several embodiments, a myriad of changes, variations, alterations, transformations, and modifications may be suggested to one skilled in the art, and it is intended that the present invention encompass such changes, variations, alterations, transformations, and modifications as fall within the spirit and scope of the appended claims.

WHAT IS CLAIMED IS:

1. A video surveillance system, comprising:

a client operable to generate data associated with a financial transaction, the client having a camera operable to generate video associated with the financial transaction, the client operable to store data and video in a digital file; and

a server coupled to the client using a communications network, the server operable to receive the digital file from the client and to store the digital file in a memory, the server having a graphical interface operable to retrieve data and video from the digital file stored in the memory for presentation.

2. The system of Claim 1, wherein the client comprises a point-of-sale device and the financial transaction comprises the sale of an item.

3. The system of Claim 1, wherein the client comprises an automated teller machine and the financial transaction comprises a cash withdrawal.

4. The system of Claim 1, wherein the client comprises a microphone operable to generate audio associated with the financial transaction, the client operable to store data, video, and audio in the digital file.

5. The system of Claim 1, wherein the digital file comprises a single multimedia digital file.

6. The system of Claim 1, wherein:

the graphical interface at the server presents data as an entry in a table having a plurality of entries associated with a plurality of financial transactions; and

the graphical interface at the server presents video in response to a selection of the entry in the table.

7. The system of Claim 1, wherein the graphical interface at the server presents data as an entry in a table having a plurality of entries associated with a plurality of financial transactions, at least one entry highlighted to indicate the existence of video associated with the entry.

8. The system of Claim 1, wherein:

the client comprises a microphone, a speaker, and a display;

the server comprises a camera, a microphone, a speaker, and a display; and

the client and the server are further operable to conduct two-way audio/video conferencing.

9. The system of Claim 1, wherein the client has a first mode and a second mode of operation, the client in the first mode includes data in the digital file, the client in the second mode includes data and video in the digital file, the second mode associated with an exception condition of the financial transaction.

10. The system of Claim 9, wherein the server transmits information to the client to define the exception condition.

11. A video surveillance method, comprising:

generating data associated with a financial transaction;

generating video associated with the financial transaction;

storing data and video in a digital file at a client;

receiving the digital file at a server using a communications network;

storing the digital file at the server; and

presenting data and video in the digital file using a graphical interface.

12. The method of Claim 11, wherein the client comprises a point-of-sale device and the financial transaction comprises the sale of an item.

13. The method of Claim 11, wherein the client comprises an automated teller machine and the financial transaction comprises a cash withdrawal.

14. The method of Claim 11, further comprising the steps of:

generating audio associated with the financial transaction; and

storing the audio in the digital file at the client.

15. The method of Claim 11, wherein the digital file comprises a single multimedia digital file.

16. The method of Claim 11, wherein the step of presenting comprises:

presenting data as an entry in a table having a plurality of entries associated with a plurality of financial transactions;

receiving a selection of the entry in the table; and

presenting video associated with the data in response to the selection of the entry in the table.

17. The method of Claim 11, wherein the step of presenting comprises:

presenting data as an entry in a table having a plurality of entries associated with a plurality of financial transactions;

highlighting the entry to indicate the existence of video associated with the data; and

presenting video associated with the data in response to the selection of the highlighted entry.

18. The method of Claim 11, wherein the step of storing data and video at the client comprises:

storing data in a first mode; and

storing data and video in a second mode associated with an exception condition of the financial transaction.

19. The method of Claim 18, further comprising the step of communicating information from the server to the client to define the exception condition of the financial transaction.

20. The method of Claim 18, wherein:

the client comprises a point-of-sale device and the financial transaction comprises the sale of an item; and

the exception condition comprises the activation of one of a selected no sale and void keys on the point-of-sale device.

21. A video surveillance system, comprising:

a client operable to generate data associated with at least one financial transaction, the client having a camera operable to generate video associated with the financial transaction, the client operable to transmit the data and video using a communications network; and

a server coupled to the client using the communications network, the server operable to receive the data and video from the client and to display the video and data in real-time.

22. The system of Claim 21, wherein the client comprises a point-of-sale device and the financial transaction comprises the sale of an item.

23. The system of Claim 21, wherein the client comprises an automated teller machine and the financial transaction comprises a cash withdrawal.

24. The system of Claim 21, wherein the client comprises a microphone operable to generate audio associated with the financial transaction, the client operable to transmit data, video, and audio over the communications network.

25. The system of Claim 21, wherein the server forms a data window from the data and a video window from the video and overlays the data window on the video window.

26. The system of Claim 21, wherein the server presents data from a plurality of financial transactions as a plurality of data windows, presents video from a plurality of video sources as a plurality of video windows, and associates the data windows with the corresponding video windows.

27. The system of Claim 26, wherein the server receives user input to specify one of the data windows to display the video window associated with the specified data window.

28. The system of Claim 26, wherein the server associated with the financial transaction automatically switches the video window to the video associated with the data in response to the presence or content of data.

29. The system of Claim 26, wherein the server displays the appropriate video window and data window upon changes in one of the plurality of video windows.

30. The system of Claim 21, wherein the client stores accumulated data associated with the financial transaction and transmits the data when the client communicates with the server.

31. A video surveillance method, comprising:

generating data associated with a financial transaction;

generating video associated with the financial transaction;

transmitting data and video in real-time from a client;

receiving the data and video at a server using a communications network; and

presenting data and video on a display at the server.

32. The method of Claim 31, wherein the client comprises a point-of-sale device and the financial transaction comprises the sale of an item.

33. The method of Claim 31, wherein the client comprises an automated teller machine and the financial transaction comprises a cash withdrawal.

34. The method of Claim 31, further comprising the steps of:

generating audio associated with the financial transaction; and

transmitting the audio to the server.

35. The method of Claim 31, wherein the step of presenting comprises:

presenting data in a data window as a representation of the financial transaction;

presenting video in a video window; and

overlaying the data window on the video window.

36. The method of Claim 31, wherein the step of presenting comprises:

presenting data as a plurality of data windows associated with a plurality of financial transactions;

presenting video as a plurality of video windows associated with a plurality of video sources; and

associating the data window with the corresponding video window.

37. The method of Claim 36, further comprising the step of updating the video window and the data window in response to the presence or content of the data in one of the plurality of data windows.

38. The method of Claim 36, further comprising the step of updating the video window and the data window in response to a change in one of the plurality of video windows.

39. The method of Claim 36, further comprising the steps of:

receiving a user selection; and

updating the video window and the data window in response to the selection.

40. The method of Claim 31, further comprising the steps of:

storing accumulated financial data in a file at the client;

transmitting the file from the client to the server upon connection of the client to the server.

41. The method of Claim 30, wherein the digital file contains financial records accumulated since last connection.

42. A video surveillance system, comprising:

a client operable to generate data associated with a financial transaction and to accumulate and store the data as a digital file, the client having a camera operable to generate video associated with the financial transaction, the client operable to transmit the data and video across a communications network; and

a server coupled to the client using the communications network, the server operable to receive the digital file upon connection with the client, to receive the data and video from the client and to display the video and data in real-time.

43. The system of Claim 42, wherein the client comprises a point-of-sale device and the financial transaction comprises the sale of an item.

44. The system of Claim 42, wherein the client comprises an automated teller machine and the financial transaction comprises a cash withdrawal.

45. The system of Claim 42, wherein the client comprises a microphone operable to generate audio associated with the financial transaction, the client operable to transmit data, video, and audio over the communications network.

46. The system of Claim 42, wherein the server forms a data window from the data and a video window from the video and overlays the data window on the video window.

47. The system of Claim 42, wherein the server presents data from a plurality of financial transactions as a plurality of data windows, presents video from a plurality of video sources as a plurality of video windows, and associates the data windows with the corresponding video windows.

48. The system of Claim 47, wherein the server receives user input to specify one of the data windows to display the video window associated with the specified data window.

49. The system of Claim 47, wherein the server associated with the financial transaction automatically switches the video window to the video associated with the data in response to the presence or content of data.

50. The system of Claim 47, wherein the server displays the appropriate video window and data window upon changes in one of the plurality of video windows.

51. The system of Claim 42, wherein the client stores accumulated data associated with the financial transaction and transmits the data when the client communicates with the server.

52. The system of Claim 42, wherein the server displays the digital file based on a configuration file.

53. A video surveillance method, comprising:

generating data associated with a financial transaction;

generating video associated with the financial transaction;

storing accumulated data as a digital file;

transmitting the digital file upon connection of the client and the server;

transmitting data and video in real-time from a client;

receiving the data and video at a server using a communications network; and

presenting data and video on a display at the server.

54. The method of Claim 53, wherein the client comprises a point-of-sale device and the financial transaction comprises the sale of an item.

55. The method of Claim 53, wherein the client comprises an automated teller machine and the financial transaction comprises a cash withdrawal.

56. The method of Claim 53, further comprising the steps of:

generating audio associated with the financial transaction; and

transmitting the audio to the server.

57. The method of Claim 53, wherein the step of presenting comprises:

presenting data in a data window as a representation of the financial transaction;

presenting video in a video window; and

overlaying the data window on the video window.

58. The method of Claim 53, wherein the step of presenting comprises:

presenting data as a plurality of data windows associated with a plurality of financial transactions on a display at the server;

presenting video as a plurality of video windows associated with a plurality of video sources on a display at the server; and

associating the data window with the corresponding video window.

59. The method of Claim 58, further comprising the step of updating the video window and the data window in response to the presence or content of the data in one of the plurality of data windows.

60. The method of Claim 58, further comprising the step of updating the video window and the data window in response to a change in one of the plurality of video windows.

61. The method of Claim 58, further comprising the steps of:

receiving a user selection; and

updating the video window and the data window in response to the selection.

62. The method of Claim 53, wherein the digital file contains financial records accumulated since last connection.

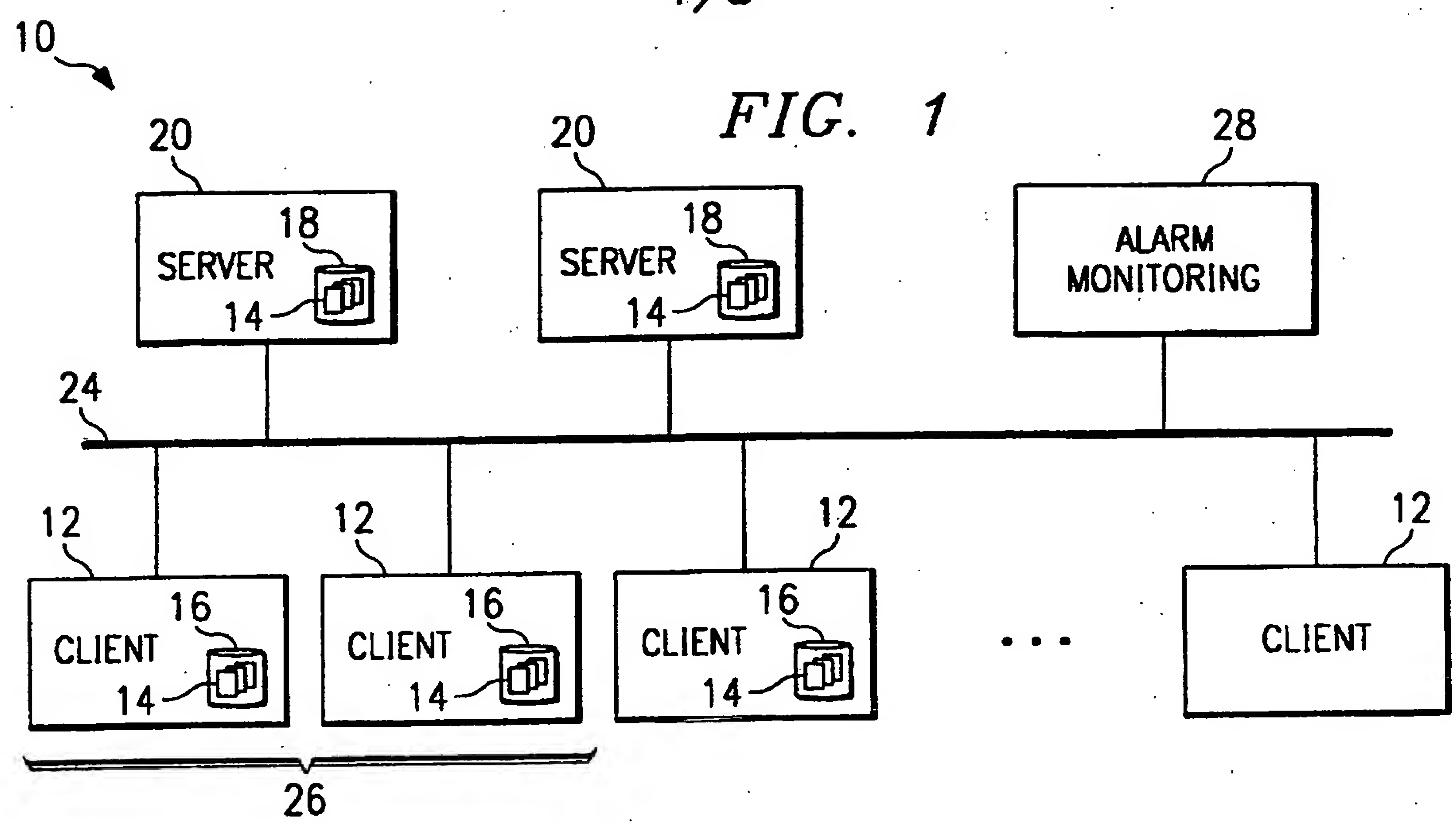

10
20
SERVER
18
14
20
SERVER
18
14
28
ALARM
MONITORING
24
12
CLIENT
16
14
12
CLIENT
16
14
12
CLIENT
16
14
12
CLIENT
26

FIG. 1

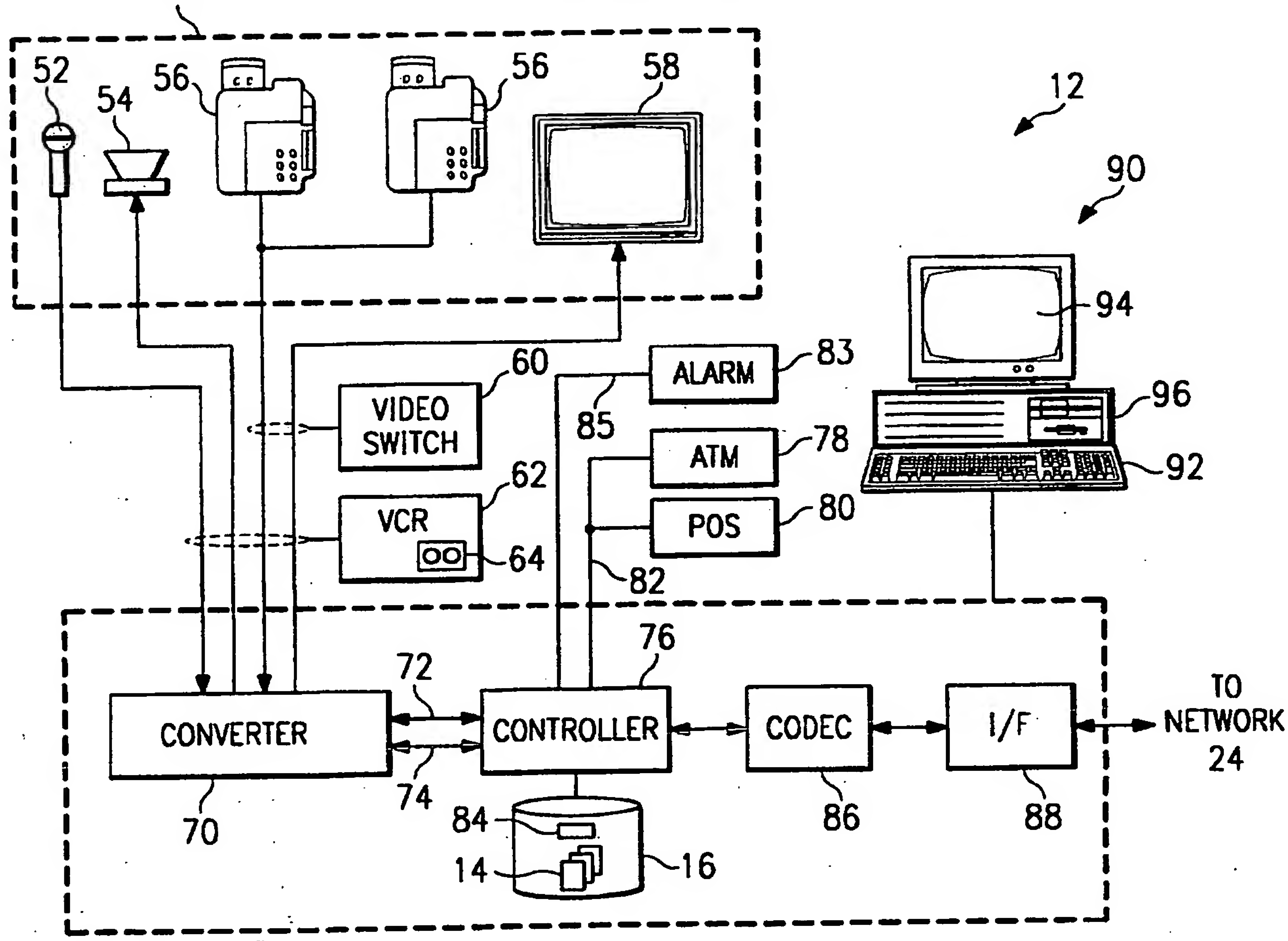

50
52
54
56
56
58
12
90
94
60
VIDEO
SWITCH
ALARM
83
85
ATM
78
96
92
62
VCR
64
POS
80
82
76
72
CONVERTER
CONTROLLER
CODEC
I/F
TO
NETWORK
24
70
74
84
14
16
86
88

FIG. 2

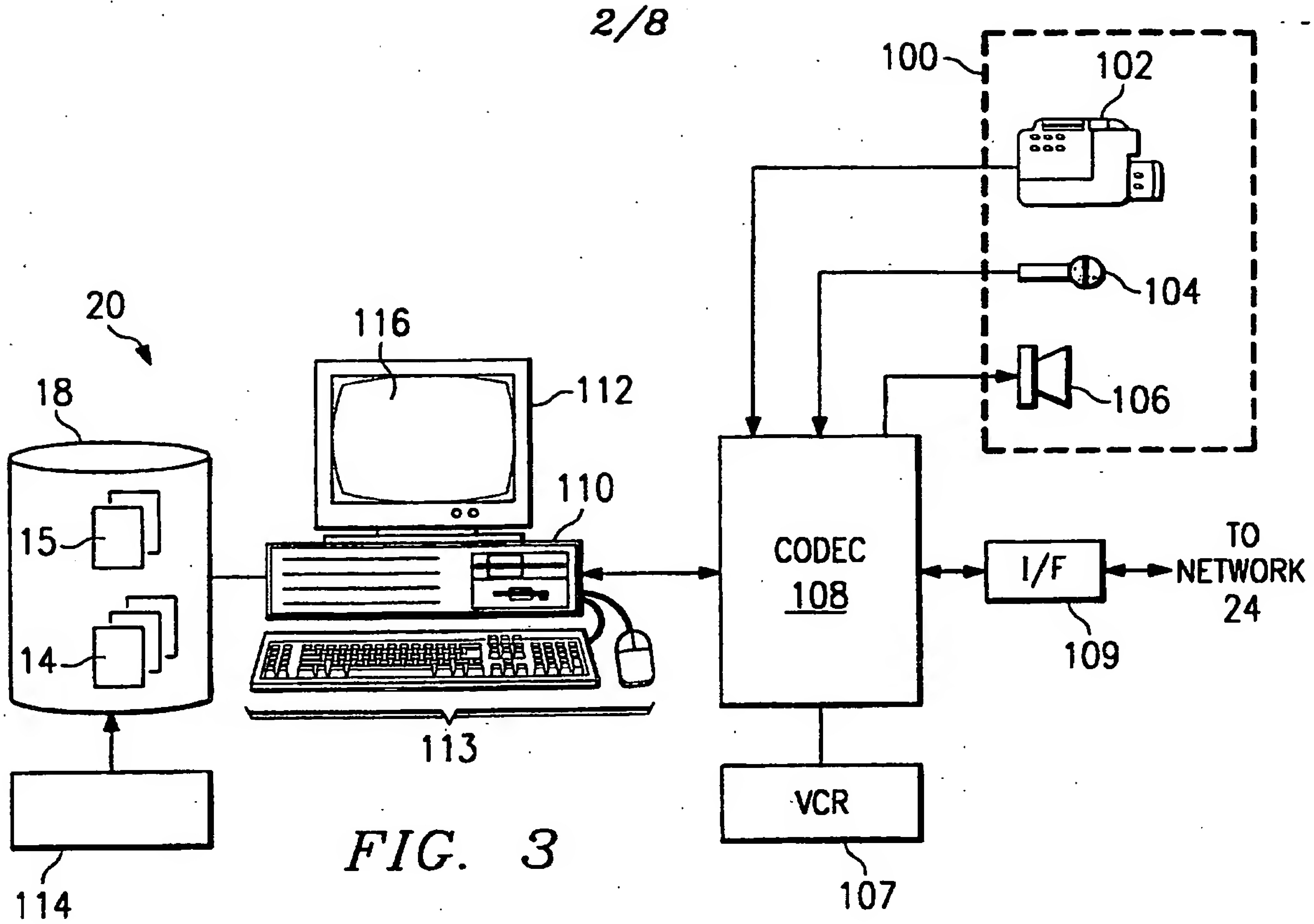

100
102
104
106
20
116
112
18
110
15
14
CODEC
108
I/F
TO
NETWORK
24
109
113
VCR
114
107

FIG. 3

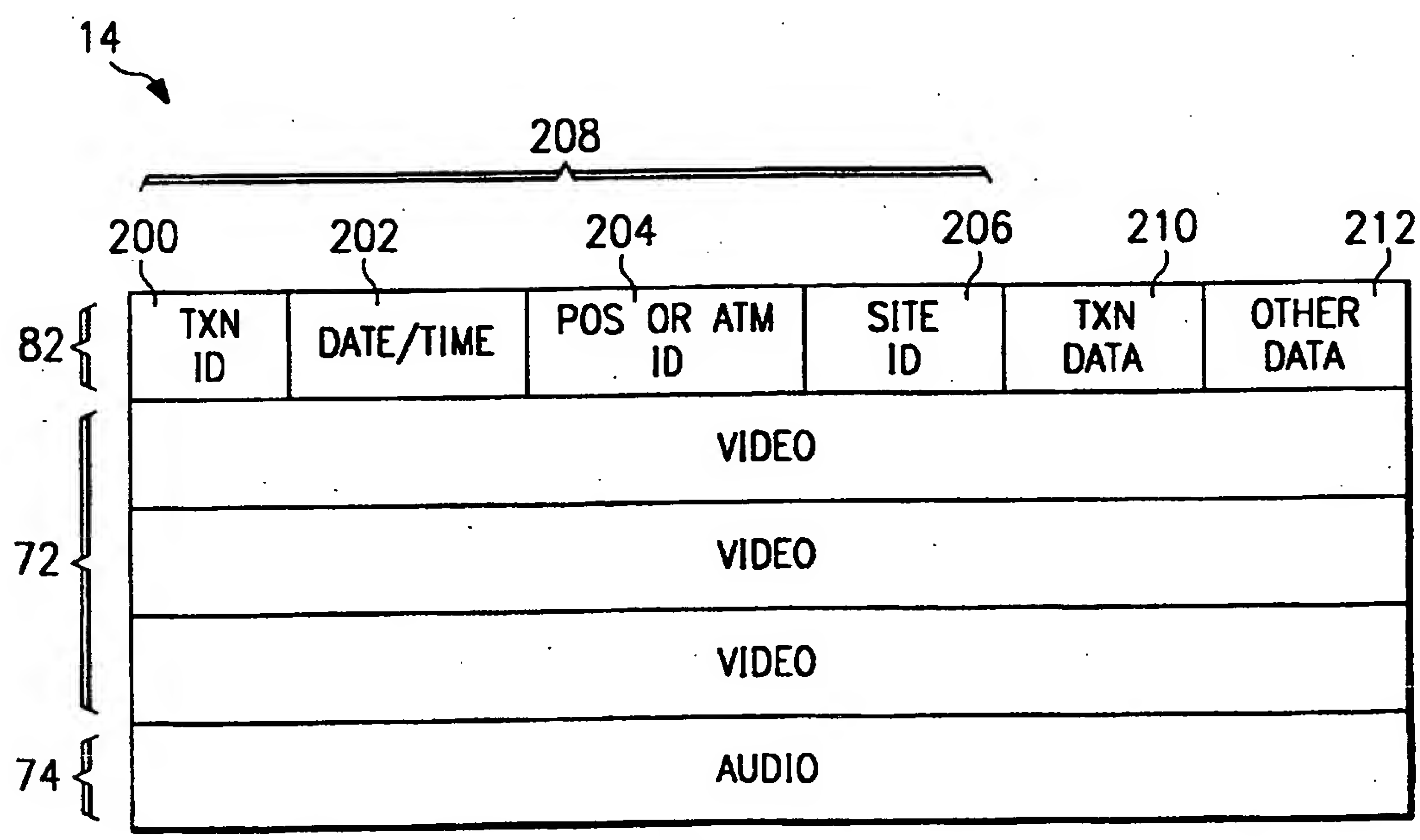
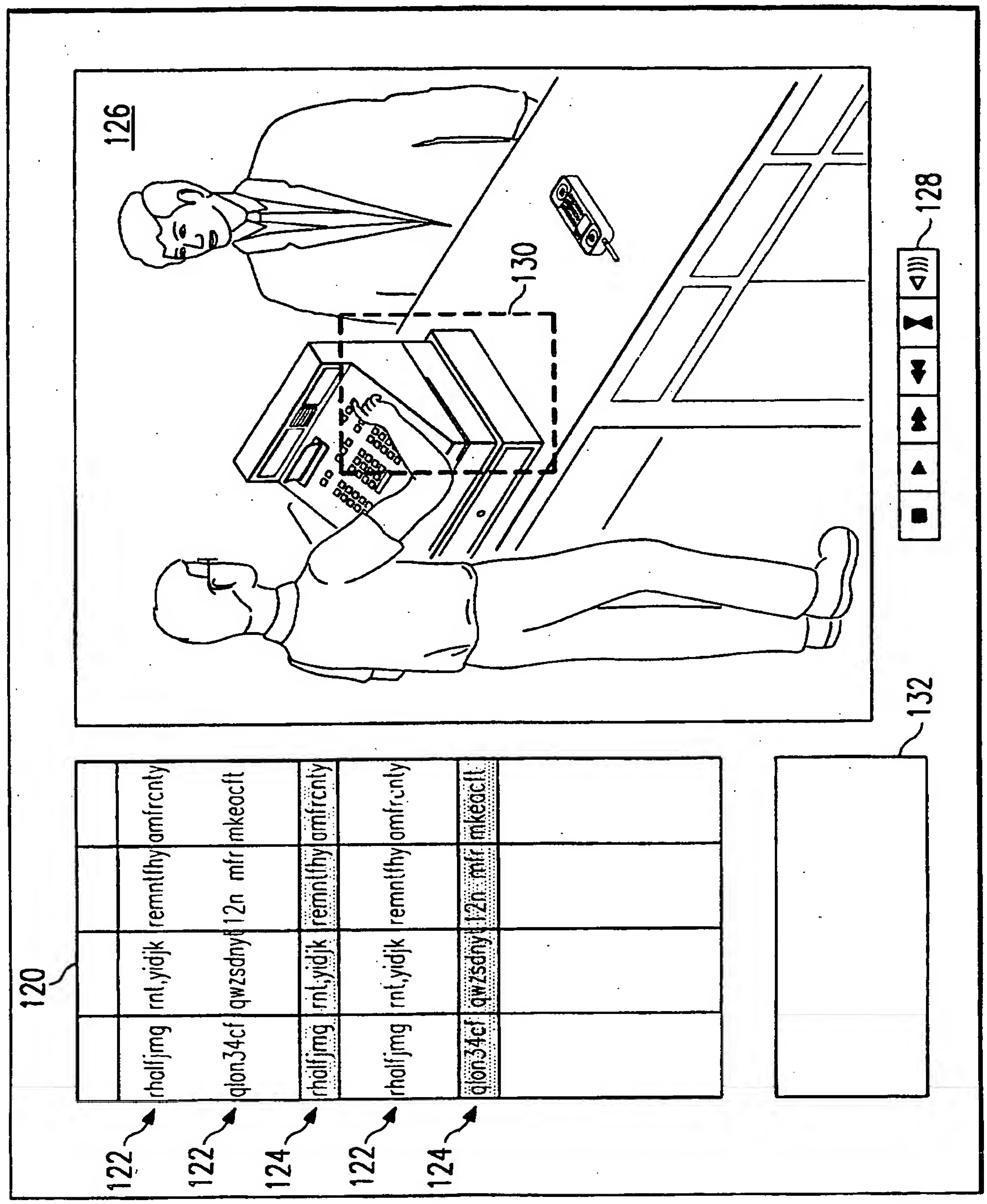

14
208
200
202
204
206
210
212
82
TXN ID
DATE/TIME
POS OR ATM ID
SITE ID
TXN DATA
OTHER DATA
VIDEO
72
VIDEO
VIDEO
74
AUDIO

FIG. 5

FIG. 4



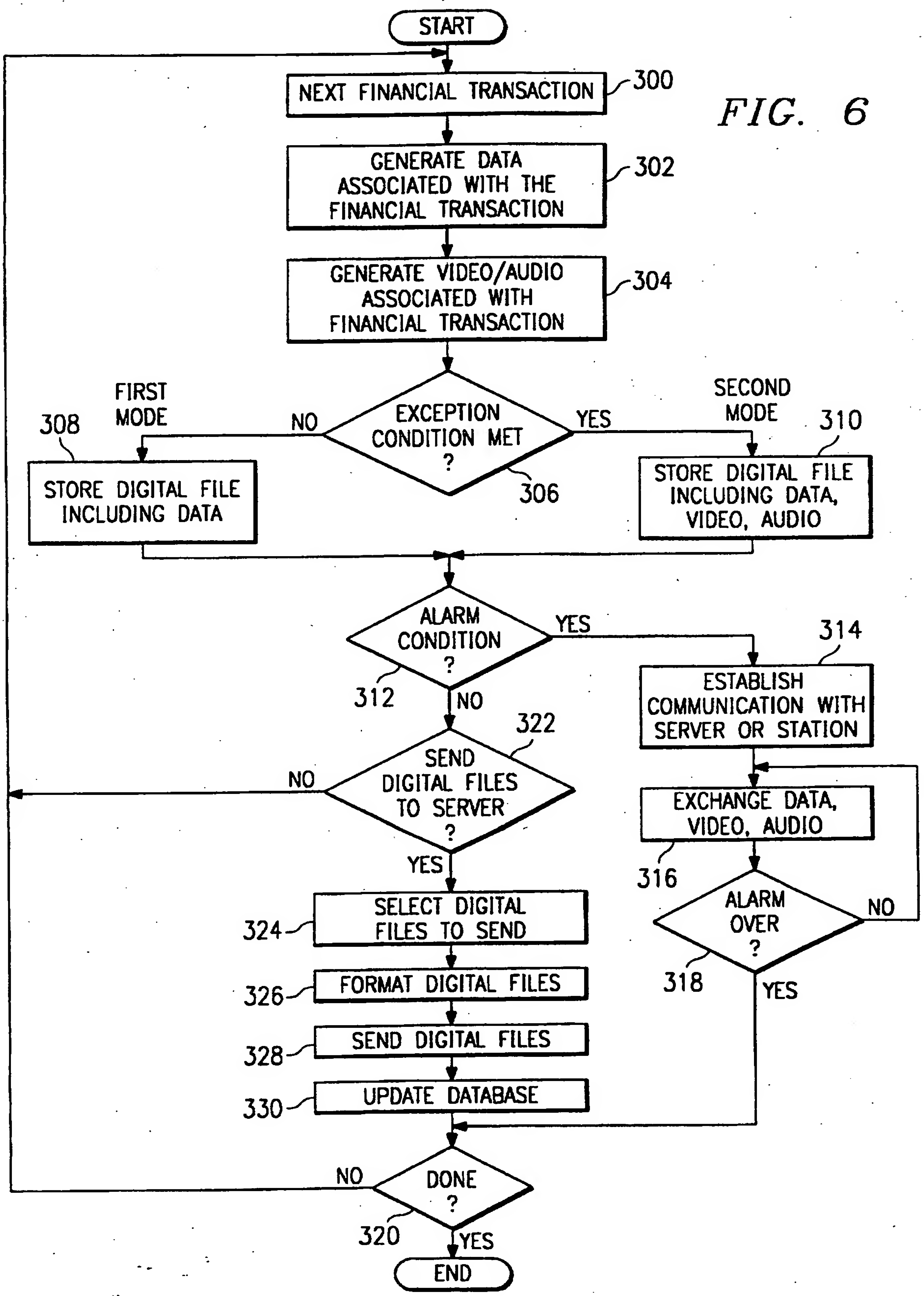

START
NEXT FINANCIAL TRANSACTION
300
GENERATE DATA ASSOCIATED WITH THE FINANCIAL TRANSACTION
302
GENERATE VIDEO/AUDIO ASSOCIATED WITH FINANCIAL TRANSACTION
304
EXCEPTION CONDITION MET ?
306
NO
FIRST MODE
308
STORE DIGITAL FILE INCLUDING DATA
YES
SECOND MODE
310
STORE DIGITAL FILE INCLUDING DATA, VIDEO, AUDIO
ALARM CONDITION ?
312
YES
314
ESTABLISH COMMUNICATION WITH SERVER OR STATION
EXCHANGE DATA, VIDEO, AUDIO
316
ALARM OVER ?
318
NO
YES
NO
SEND DIGITAL FILES TO SERVER ?
322
NO
YES
SELECT DIGITAL FILES TO SEND
324
FORMAT DIGITAL FILES
326
SEND DIGITAL FILES
328
UPDATE DATABASE
330
DONE ?
320
NO
YES
END

FIG. 6

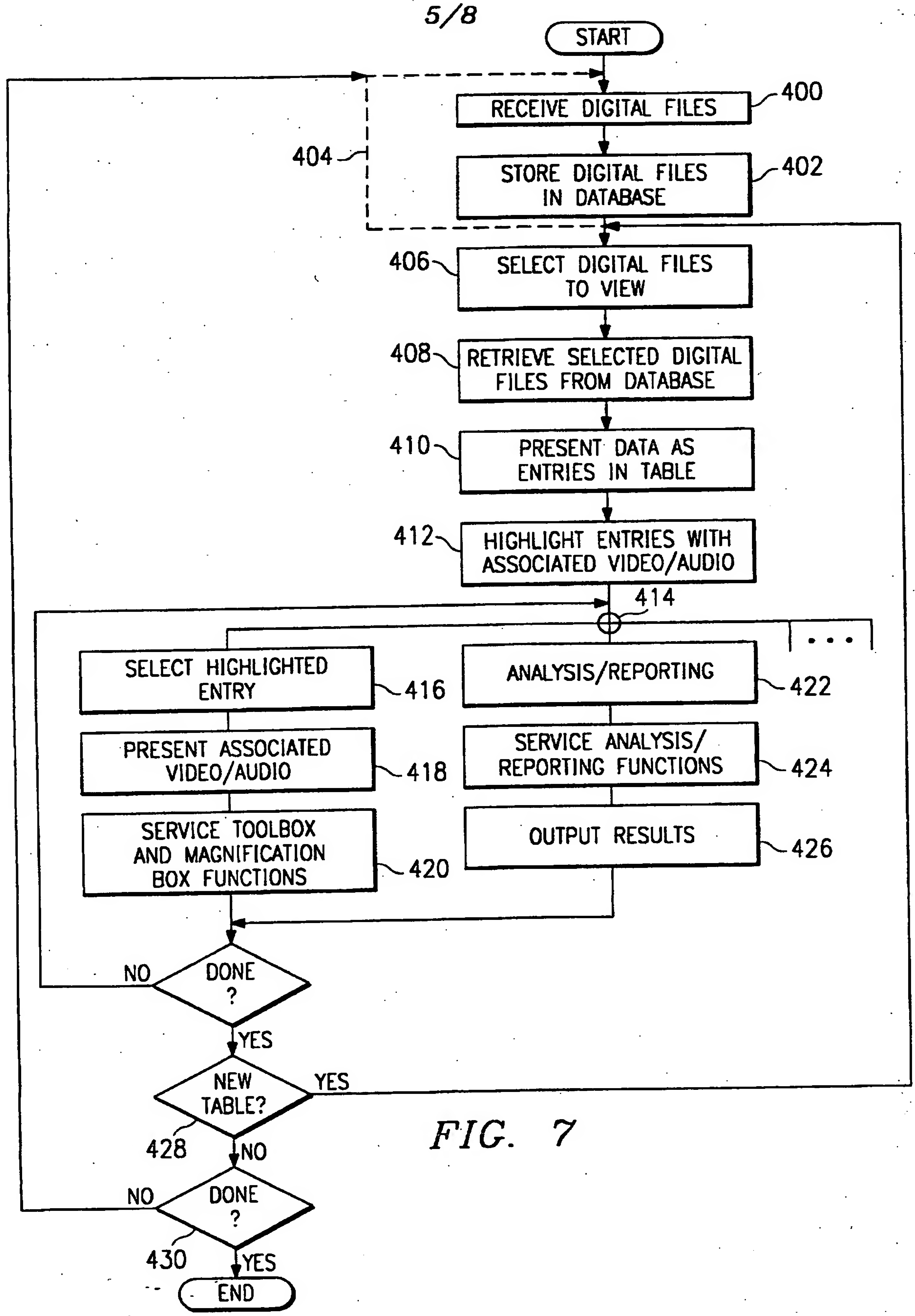

START
RECEIVE DIGITAL FILES
400
STORE DIGITAL FILES IN DATABASE
402
404
SELECT DIGITAL FILES TO VIEW
406
RETRIEVE SELECTED DIGITAL FILES FROM DATABASE
408
PRESENT DATA AS ENTRIES IN TABLE
410
HIGHLIGHT ENTRIES WITH ASSOCIATED VIDEO/AUDIO
412
414
SELECT HIGHLIGHTED ENTRY
416
PRESENT ASSOCIATED VIDEO/AUDIO
418
SERVICE TOOLBOX AND MAGNIFICATION BOX FUNCTIONS
420
ANALYSIS/REPORTING
422
SERVICE ANALYSIS/ REPORTING FUNCTIONS
424
OUTPUT RESULTS
426
• • •
DONE ?
NO
YES
NEW TABLE?
YES
428
NO
DONE ?
NO
430
YES
END

FIG. 7

FIG. 8

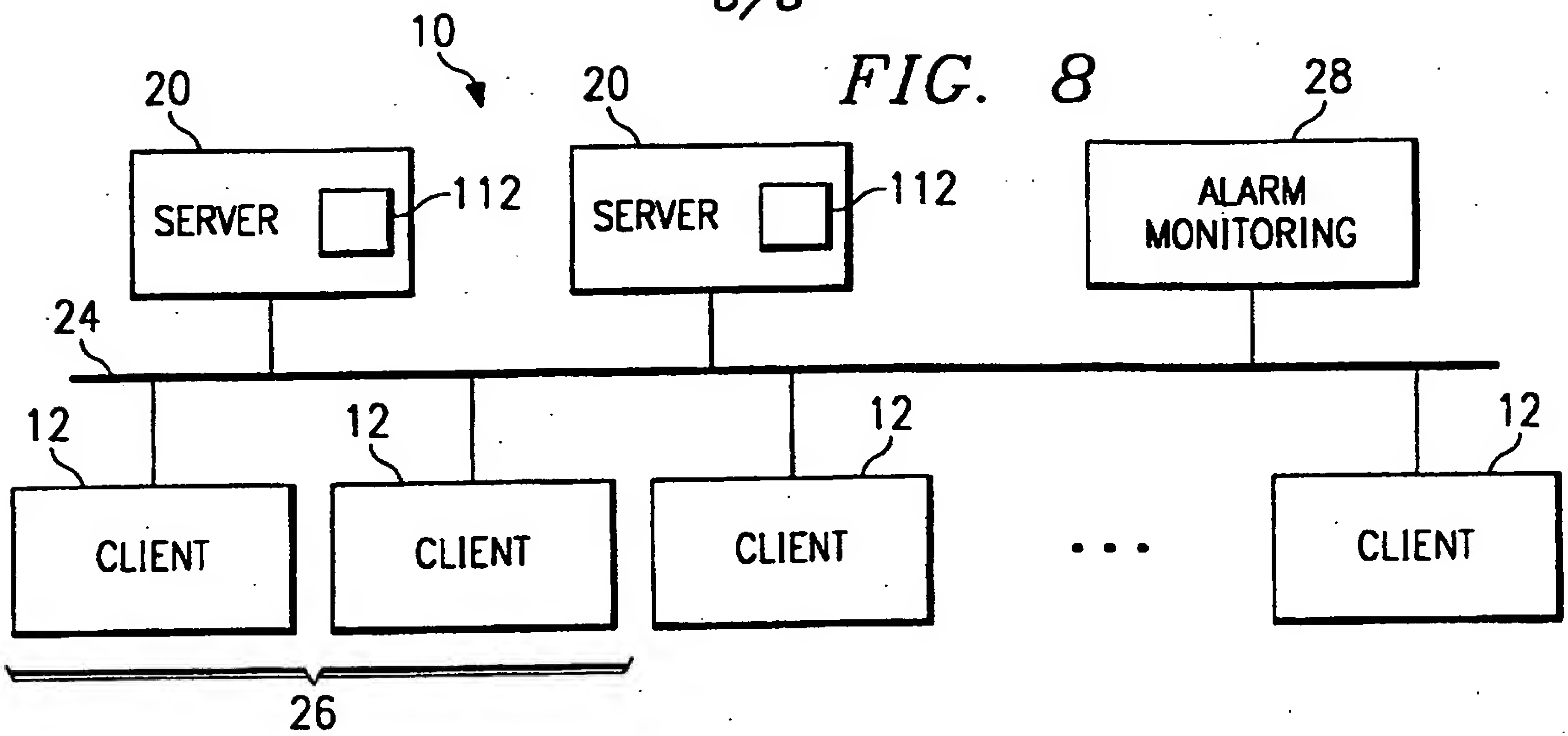
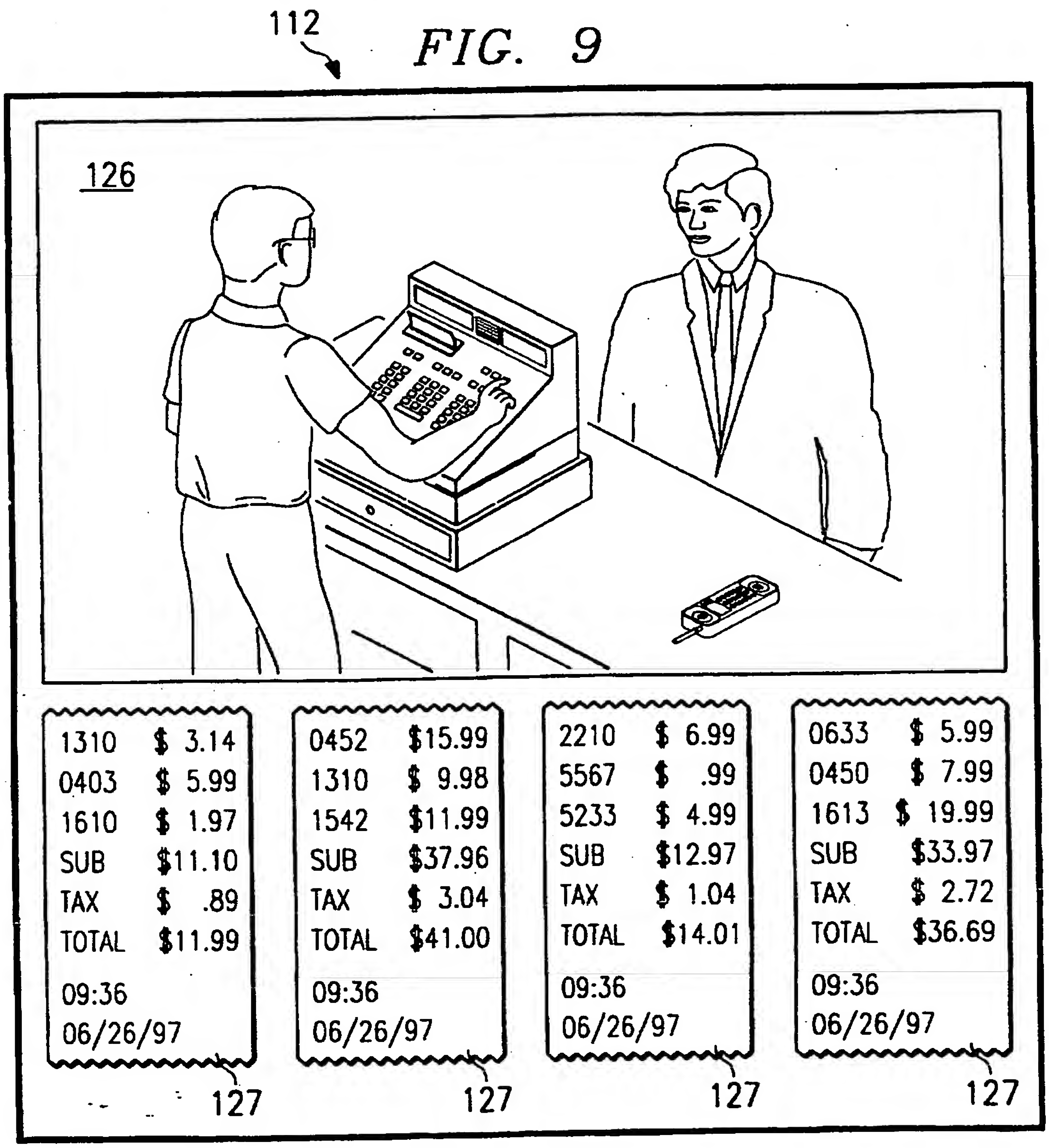

10
20
SERVER
112
20
SERVER
112
28
ALARM MONITORING
24
12
CLIENT
12
CLIENT
12
CLIENT
• • •
12
CLIENT
26

FIG. 9


112
126
1310 $ 3.14
0403 $ 5.99
1610 $ 1.97
SUB $11.10
TAX $ .89
TOTAL $11.99
09:36
06/26/97
0452 $15.99
1310 $ 9.98
1542 $11.99
SUB $37.96
TAX $ 3.04
TOTAL $41.00
09:36
06/26/97
2210 $ 6.99
5567 $ .99
5233 $ 4.99
SUB $12.97
TAX $ 1.04
TOTAL $14.01
09:36
06/26/97
0633 $ 5.99
0450 $ 7.99
1613 $ 19.99
SUB $33.97
TAX $ 2.72
TOTAL $36.69
09:36
06/26/97
127
127
127
127

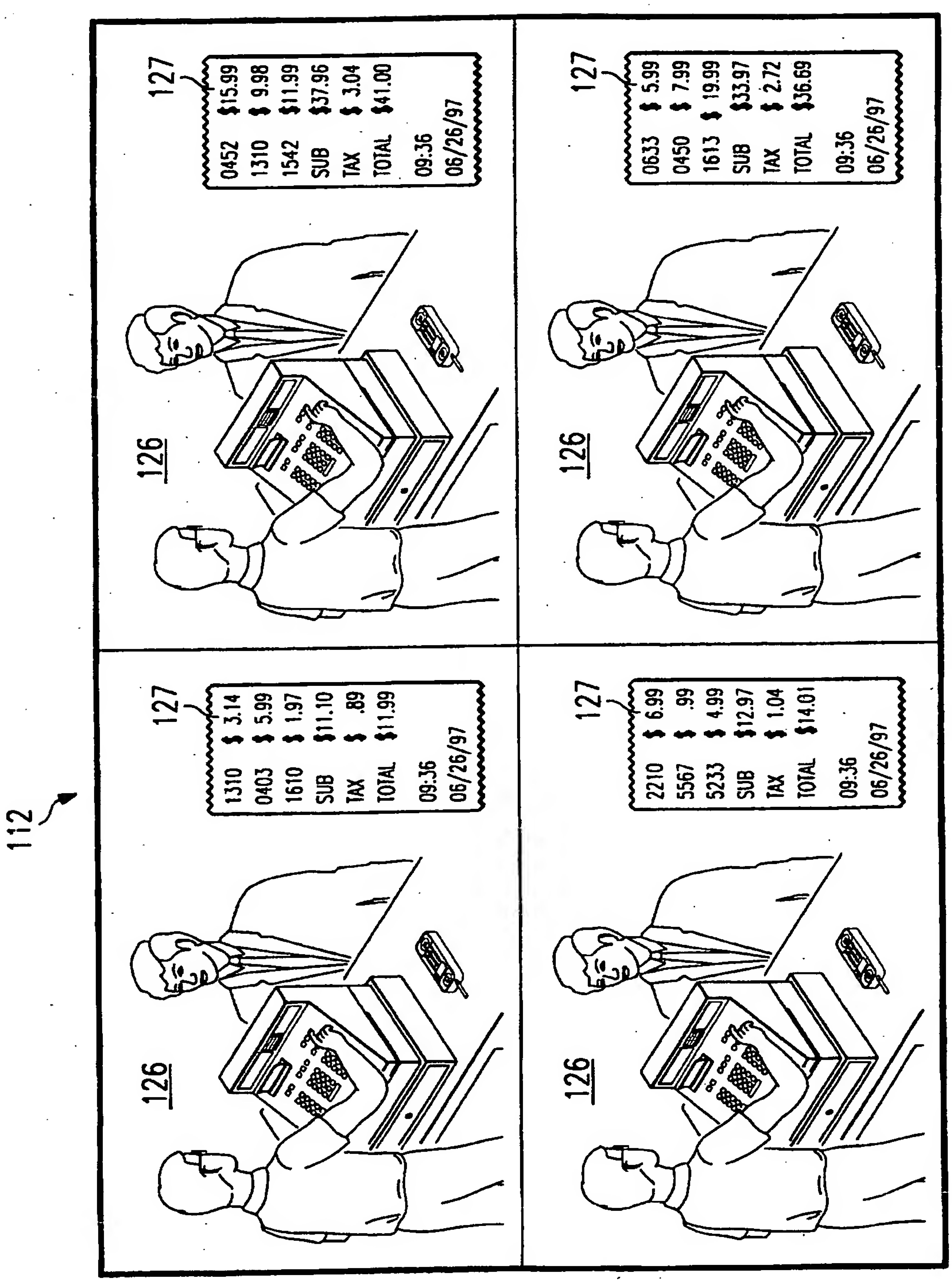

112
126
127
1310 $ 3.14
0403 $ 5.99
1610 $ 1.97
SUB $11.10
TAX $ .89
TOTAL $11.99
09:36
06/26/97
126
127
0452 $15.99
1310 $ 9.98
1542 $11.99
SUB $37.96
TAX $ 3.04
TOTAL $41.00
09:36
06/26/97
126
127
2210 $ 6.99
5567 $ .99
5233 $ 4.99
SUB $12.97
TAX $ 1.04
TOTAL $14.01
09:36
06/26/97
126
127
0633 $ 5.99
0450 $ 7.99
1613 $ 19.99
SUB $33.97
TAX $ 2.72
TOTAL $36.69
09:36
06/26/97

FIG. 10

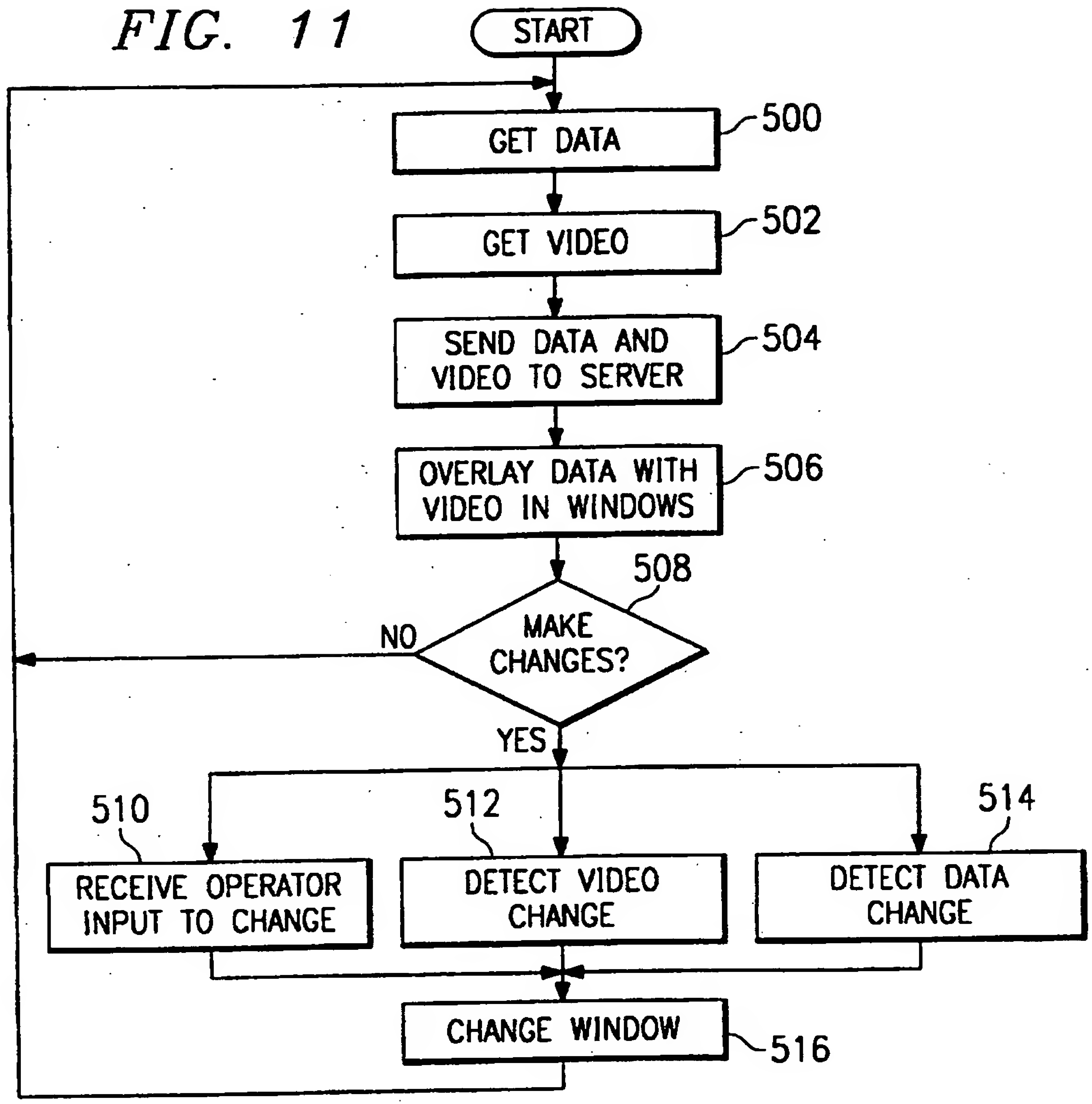

START
GET DATA
500
GET VIDEO
502
SEND DATA AND VIDEO TO SERVER
504
OVERLAY DATA WITH VIDEO IN WINDOWS
506
508
MAKE CHANGES?
NO
YES
510
512
514
RECEIVE OPERATOR INPUT TO CHANGE
DETECT VIDEO CHANGE
DETECT DATA CHANGE
CHANGE WINDOW
516

FIG. 11

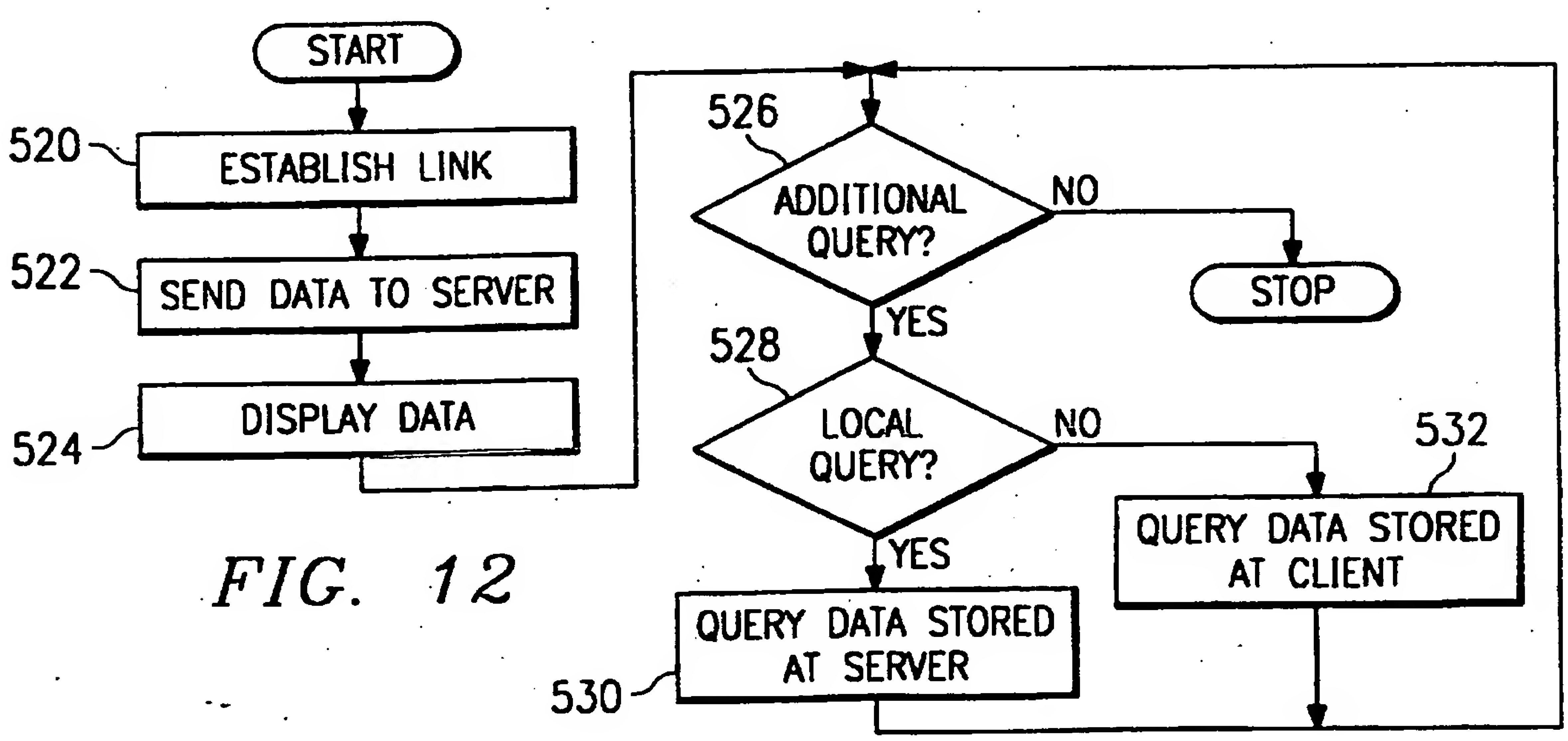

START
520
ESTABLISH LINK
522
SEND DATA TO SERVER
524
DISPLAY DATA
526
ADDITIONAL QUERY?
NO
STOP
YES
528
LOCAL QUERY?
NO
532
QUERY DATA STORED AT CLIENT
YES
QUERY DATA STORED AT SERVER
530

FIG. 12

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/12000

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G08B15/00 G07F7/10 G07G3/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G07F G07G G08B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 663 655 A (INTEGRATED SECURITY) 19 July 1995	1,2,11, 12,21, 22,25, 26,31, 32,35,36
Y	see page 10, line 42 - page 11, line 32; figures 1-4	3-5, 13-15, 23,24, 33,34, 45,56
A		6-9, 16-18, 20, 27-30, 34, 37-44, 46-55, 57-62

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
11 November 1997	05. 12. 97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Neville, D

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/12000

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 661 677 A (EASTMAN KODAK) 5 July 1995	3,13,23,33
A	see column 3, line 40 - line 57; figure 1	1,11,18,19,21,24,30,40-42,44,45,51,53,55,62

Y	US 4 054 752 A (DENNIS, JR. ET AL.) 18 October 1977	4,5,14,15,24,34,45,56
A	see column 6, line 46 - line 62; figures 1,4	1,2,8,11,12,20-22,31,32,42,43,53,54

X	US 4 991 008 A (NAMA) 5 February 1991	21-23,25,31-33,35
	see column 6, line 25 - line 66; figure 1 see column 9, line 5 - line 46; figure 3	
A	see column 11, line 32 - column 12, line 2; figure 4	1-3,8,11-13,19,20,24,30,40-46,51,53-55,57,62

A	EP 0 332 161 A (ROBOTFOTO & ELECTRONIC) 13 September 1989	1,3,9,11,13,18,21,23,30,31,33,40-42,44,51,53,55,62
	see column 2, line 45 - column 3, line 7; figure 2	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 97/12000

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 663655 A	19-07-95	US 5216502 A	01-06-93
		AT 127601 T	15-09-95
		AU 660005 B	08-06-95
		AU 9166391 A	22-07-92
		CA 2098596 A	19-06-92
		DE 69112842 D	12-10-95
		DE 69112842 T	04-04-96
		DE 563303 T	17-03-94
		EP 0563303 A	06-10-93
		ES 2077396 T	16-11-95
		JP 6503437 T	14-04-94
		OA 9781 A	15-04-94
		WO 9211614 A	09-07-92
EP 661677 A	05-07-95	JP 7234958 A	05-09-95
US 4054752 A	18-10-77	NONE	
US 4991008 A	05-02-91	NONE	
EP 332161 A	13-09-89	DE 3807727 A	21-09-89
		CA 1323086 A	12-10-93
		JP 1296398 A	29-11-89
		US 4942464 A	17-07-90

BEST AVAILABLE COPY

55

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G07G 3/00, H04N 7/18, G08B 13/196 — A1

(11) International Publication Number: WO 95/29470

(43) International Publication Date: 2 November 1995 (02.11.95)

(21) International Application Number: PCT/US95/05167

(22) International Filing Date: 24 April 1995 (24.04.95)

(30) Priority Data:
08/232,363 — 25 April 1994 (25.04.94) — US

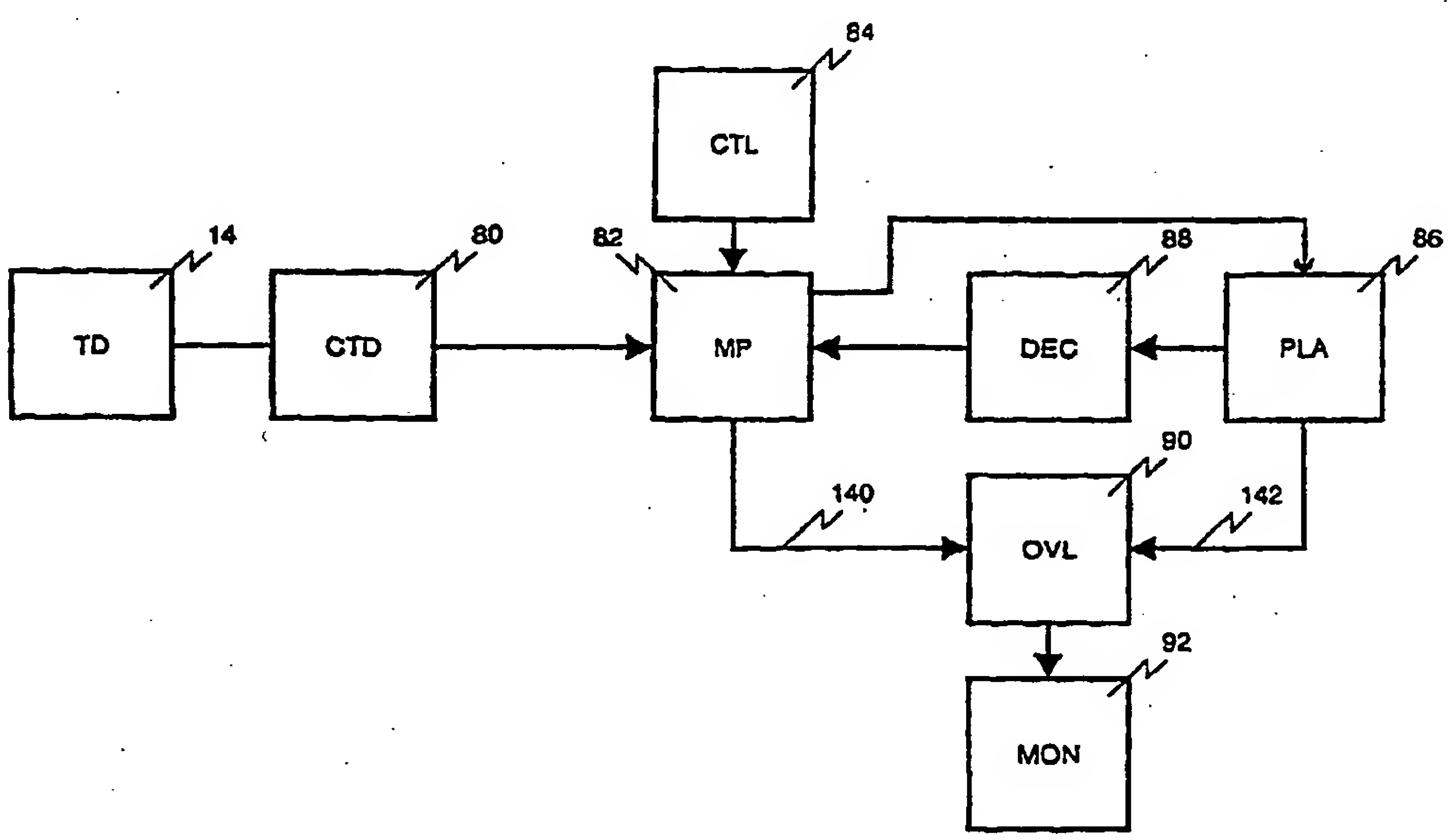
(71)(72) Applicant and Inventor: KATZ, Barry [US/US]; 503 Cindy Circle, Penllyn, PA 19422 (US).

(74) Agent: GARZIA, Mark, A.; Synnestvedt & Lechner, 2600 One Reading Center, 1101 Market Street, Philadelphia, PA 19107-2950 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.

(54) Title: ASYNCHRONOUS VIDEO EVENT AND TRANSACTION DATA MULTIPLEXING TECHNIQUE FOR SURVEILLANCE SYSTEMS



(57) Abstract

A surveillance system which can store and replay information which is not generated contemporaneously. In some point-of-sale systems, the behavioral events occur before the transaction data is generated. The present system uses an independently generated synchronizing signal which is recorded with the video signals corresponding to the behavioral events and the transaction signal corresponding to the transaction events. Upon playback, the transaction data is matched up with the behavioral event being replayed from the video tape by utilizing the synchronizing signals stored with each signal.

BEST AVAILABLE COPY

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

BEST AVAILABLE COPY

ASYNCHRONOUS VIDEO EVENT AND TRANSACTION DATA MULTIPLEXING TECHNIQUE FOR SURVEILLANCE SYSTEMS

Field of the Invention

This invention relates to surveillance systems that record transaction events for review at a later date. More specifically, this invention relates to an asynchronous video event and transaction data multiplexing technique for such a surveillance system.

Background of the Invention

The use of surveillance systems to record cash transactions for later review are well known in the art. For example, U.S. Patent No. 4,337,482, to Coutta, discloses a surveillance system that monitors and records transactions that occur at a number of cashier lanes. In Coutta, a single television camera, mounted on a rail, can be positioned to make a video recording of the transactions that occur at a single selected cashier lane. Coutta discloses that the digital transaction data from the cash register in the selected cashier lane is fed into a video character generator to provide a composite video picture in which an alphanumeric display of the transaction data overlays the video image of

the transaction. Since a composite video image is generated with respect to only one cashier lane, it is usually possible to position the camera so that the alphanumeric overlay does not obscure a useful portion of the recorded video image. However, if a single camera is used to record the transactions that occur at a plurality of cashier lanes, it is likely that the alphanumeric overlay data will obscure an important part of the video image of at least one of the transaction lanes. This likelihood is further increased when a large number of parameters are displayed simultaneously for all of the cashier lanes.

In U.S. Patent No. 4,630,110, to Cotton et al., a surveillance system is disclosed which monitors and records events from a single transaction lane by a plurality of video cameras. In one embodiment of Cotton et al., the video image from four cameras are combined, with two of the cameras being focused on the visual read-out of the cash registers. In Cotton et al., the textual data can be displayed at the lower portion of the combined video picture.

Another surveillance system disclosed in U.S. Patent No. 4,145,715, to Clever, generates two levels of surveillance records. The first level, generated by a tape recorder, contains a record of all transactions. The second level generated on the tape recorder contains only selected transactions. In Clever, transaction data, such as the price and department number, are input to a character generator.

The character generator output is mixed with the video image to create a composite video signal. This composite video signal consists of alphanumeric transaction data which overlays the transaction video image and is recorded by a video tape recorder onto a video tape.

Although Clever discloses that a single camera can be used to scan several point-of-sale (POS) terminals, the stored composite video signal on playback always contains alphanumeric transaction data that is permanently overlaid on the video image. Accordingly, upon playback of the composite video signal, a portion of the video image cannot be seen (i.e., the portion "under" the alphanumeric transaction data) and this portion can never be recovered. The alphanumeric overlay degrades the clarity of the resultant video images, especially if the transaction data is placed over the video image corresponding to the desired cashier lane (i.e., the cashier lane directly corresponding to the transaction data). Alternatively, a portion of the video may be "blacked out" so that the transaction data can be more easily read when viewed at a later time on the monitor. In this instance, the blacked out portion is recorded over a portion of the image being recorded by the television camera. Again, the portion of the video image which was blacked out is lost forever. These problems arise in the Clever system because the composite video signal is generated before recording on the video tape.

As the devices that perform data entry (cash registers, data terminals, optical character readers, radio frequency readers, magnetic media readers, etc.) become more sophisticated, larger quantities of alphanumeric characters describing the transaction are generated. The increase in information desired to be recorded for each lane would further tend to clutter and obscure the composite video image. Further, as the number of lanes being monitored increases, it becomes more difficult to overlay all of the alphanumeric transaction data at positions that will not obscure an important part of the video transaction image. Another factor to consider when recording the transaction data over the video images are the varying light and weather conditions when the transaction lanes are outdoors.

In U.S. Patent No. 5,216,502, to Katz, the transaction data and video pictures of the transaction behavior are recorded synchronously but separately on media capable of storing a full motion video.

The Katz patent (and for that matter all of the aforementioned patents) require that the transaction data be available at the time that the behavior is being recorded since the video images and the transaction data are recorded contemporaneously. However, there are certain applications where this technology cannot be applied. For example, in situations where the point-of-sale terminal buffers the transaction data until the termination of the transaction, or

at the termination of several transactions. At the end of the transaction, the data is transmitted from the POS terminal to a host computer (i.e., the data is sent out in one burst instead of as a continuous stream). Accordingly, the transaction information cannot be recorded synchronously with the video images of the transaction.

Summary of the Invention

It is an object of this invention to provide an improved surveillance system.

It is a further object to provide an asynchronous surveillance system.

A surveillance system for reviewing behavioral and transaction events which occur at one or more POS terminals, check outs, transaction lanes or operation stations is disclosed. The subject surveillance system includes means for generating video image signals, means for generating transaction signals, means for generating a synchronizing signal, means for associating the synchronizing signal with the video signal and the transaction signal, and a means for recording the video signal and transaction signal along with their associated synchronizing signals. The surveillance system also includes a means for recovering the recorded signals, means for utilizing the synchronizing signal to

synchronize the video signal with its corresponding transaction signal, means for generating a composite video signal which consists of overlaying an alphanumeric representation of at least a portion of the transaction signal over its corresponding video signal, and means for displaying the composite video signal.

Light recordings of the behavioral events occurring at a POS terminal are made by a video or television camera, or any device that records light/images. The behavioral events include the customer's actions and the cashier's actions occurring at the POS terminals. The video camera senses the behavioral events and generates video signals corresponding to these events. A first recording device then records the video signal. This recording is preferably made in real time. Accordingly, a behavioral history at each transaction lane is made.

The first recording device is usually a video tape recorder or VCR. The video signals generated by the video camera are then stored on a video tape.

A sensor means at the transaction lane senses the transaction events occurring at the POS terminal and generates a transaction signal corresponding to the transaction events. The sensor means can be a cash register, a toll booth register, a machine that automatically receives money at toll plazas, a machine that can read a bar code printed on an item

(bar code scanners), or any other point-of-sale device. The transaction events include the purchase of items or goods in a store, the payment of a toll, etc. A second recording device, preferably a host computer, connected to the sensor means, records the transaction signal and records it in a database. Accordingly, a transaction history of each terminal is made.

The transaction signals generated by the sensor means are usually in a digital format and are sometimes referred to as digital signals. The signals from the sensor means may include a variety of information besides the item name and cost. For example, the sensor signals may include any transaction lane identifier, time, date, camera identification and data source identification.

The second recording means stores the digital signals along with all other event or transaction data on a second recording medium. If the second recording device is a computer, the second recording medium is usually a floppy disk.

A means for generating a synchronizing signal is required. The synchronizing signal is usually a clock signal which can be generated by an independent clock or from the host computer's clock. The synchronizing signal can also be a transaction sequence number generated by the point-of-sale terminal or any other convenient signal which can be used as a reference. The synchronizing signal is then added to both the

video signal and the transaction signal before they are recorded. Therefore, the video tape and the floppy disk contain the synchronizing signal which is used to match up the behavioral history with the corresponding transaction history.

Depending on the media used to store the video signals and the transaction signals, playback means are required to retrieve both signals along with the synchronizing information. If the first and second recording means use different storage media, two different playback means may be required. In the preferred embodiment a VCR is used to retrieve the video signals from the videotape and a computer is used to retrieve the transaction signal from the floppy disk.

A control and processing means (which can be the same computer used to retrieve the transaction signal) synchronizes the transaction signal from the second recording medium with the video signal from the first recording medium by comparing the synchronizing signals stored on both recording media. In this manner, the processing means is able to call up the transaction history corresponding to the exact behavioral history being played by the playback device. The processing means then generates a video overlay signal which includes data for an alphanumeric representation of the transaction signal. An overlay generator generates a composite video signal from the playback video signal and the video overlay signal. The composite video signal includes information

representing an alphanumeric display of the transaction history overlayed on the corresponding behavioral history. A monitor is used to display the composite video signal.

The AVETDM surveillance system can be customized for a particular application. The control and processing means includes a means for setting parameters in response to an input signal from an operator. Some of the more common features allow the operator to select a desired POS terminal, to select all transactions involving credit cards, to select all transactons involving tractor trailers, to select all transactions in which the cashier took over five minutes to ring out a customer, to select all transactions in which a customer redeemed coupons, etc.

An alternate embodiment may use a single recording device to store both the video signals and the transaction signals onto a single recording medium. If the transaction signal is recorded on the same medium as the video signal, it should be stored separately so as not to degrade either the video signals or the transaction signals. For example, if the recording medium is a video tape, the transaction signal may be stored on the audio portion.

Another embodiment would store the transaction data on the video tape even though a separate recording means for the transaction signal is employed. Therefore, the transaction signal is recorded twice - once on the video tape and once on

the floppy disk. This duplicate recording is required when the surveillance system is to be used as verifiable evidence in a judicial proceeding (e.g., to prove that a cashier was stealing). The presence of the transaction signal on the video tape is used as a sequential record to refute any charge that the transaction data on the floppy disk was altered. However, the AVETDM system does not normally use the transaction signal stored on the video tape.

Brief Description of the Drawings

These and other objects of the present invention and the various features and details of the operation and construction thereof are hereinafter more fully set forth with reference to the accompanying drawings, where:

Fig. 1 is a schematic block diagram of a surveillance and transaction recording system in accordance with the instant invention.

Fig. 2 is a schematic block diagram of an alternative system for receiving time synchronization information for use in a surveillance system in accordance with the present invention.

Fig. 3 is a schematic block diagram of an alternative system for recording the behavioral events in accordance with the present invention.

Fig. 4 is a schematic block diagram of a system for recording transactions continuously for an extended period of time.

Fig. 5 is a schematic block diagram of an alternative system for recording transactions from a plurality of transaction lanes continuously for an extended period of time.

Fig. 6 is a schematic block diagram of the interface configuration.

Fig. 7 is a block diagram of an alternative configuration of the interface.

Fig. 8 is a block diagram of an alternative system for recording transactions for an extended period of time which may be used with the system shown in Fig. 3.

Fig. 9 is a block diagram of the instant invention using a common time source.

Fig. 10 is a block diagram of an alternative system using a common time source.

Fig. 11 is a block diagram of a second alternative system using a common time source.

Fig. 12 is a replay station used to review the transactions.

Fig. 13 is a schematic block diagram of a surveillance and transaction recording system for use with a multiple point-of-sale terminal in accordance with the present invention.

Detailed Description of the Preferred Embodiment

Most point-of-sale systems utilize a plurality of cash registers which are all connected to a central processing means or host computer. The format used by manufacturers of cash registers and other data entry terminals for transmitting and storing transaction information is extremely diverse. Some systems transmit the transaction data to the host computer after every item is entered or "rung up". (These systems are sometimes referred to as "continuous stream systems.") Other point-of-sale systems store the transaction data in the cash registers until all items from one customer are entered and the transaction is consummated (i.e., after the total is determined), or after all items from several customers are entered. The transaction data representing the purchase of a number of items is then sent as one bundle or

burst of information to the central processing means for permanent storage in a database. (These systems are sometimes referred to as "burst systems".)

Previous surveillance systems are well-adapted to record information from continuous stream systems. These previous surveillance systems would tap into the wires or bus connecting the cash register to the host computer, and record the transaction signal contemporaneously with the behavioral events recorded by a television camera. One of the more advanced surveillance systems of this type was disclosed in U.S. Application No. 07/629,255, filed December 18, 1990, which issued on June 1, 1993, as U.S. Patent No. 5,216,502. U.S. Patent 5,216,502 is incorporated by reference as if fully set forth herein. However, these previous surveillance systems can not be used for burst-type point-of-sale systems since the transaction signal is not generated until some time after the behavioral events are detected and recorded. During playback, the previous surveillance systems cannot synchronize the video signals with the later generated transaction signals.

Fig. 1 is a block diagram of the preferred embodiment of the recording system of the Asynchronous Video Event and Transaction Data Multiplexing (AVETDM) surveillance/recording system in accordance with the present invention. The surveillance system can be used in different point-of-sale (POS) environments such as a retail environment (grocery

stores, convenience stores, music/specialty stores, etc.), automotive vehicle toll booths employed at turnpikes, tunnels and bridges, and electronic guard tours. It is especially well suited for point-of-sale terminals which delay the transmission of the transaction information until the completion of a customer transaction or after several customer transactions.

The recording system of the present invention is generally indicated at 10. The transaction events (e.g., the purchase of goods) occur at a point-of-sale (POS) station or terminal 12. The POS terminal 12 can be an automatic teller machine, cash register, bar code scanner, toll booth or other system. As goods are recorded or rung up by the POS terminal 12, it temporarily stores the data in a buffer. At least the name and price of every item purchased by a customer is stored. At the end of the transaction, the transaction data is forwarded to transaction database means 14. The stored transaction data corresponds to a transaction history of the goods purchased. The transaction database means 14 can be a central processing unit or host computer with the transaction data stored on a magnetic medium (e.g., a floppy disk).

A camera 16 is positioned so that it can view the customer's and/or cashier's behavior at one or more POS terminals 12. The camera 16 generates a video signal which is stored on a recording means 22. The stored video signal

corresponds to a behavioral history at the time the goods are purchased.

Since the transaction signal is generated and stored at a point in time after the video signal has been stored, the subject AVETDM surveillance system requires a synchronizing signal in order to align the transaction signal with its corresponding video signal upon playback. In the preferred embodiment, the synchronizing signal is generated by a common time source 18 which provides the synchronizing information for the transaction database 14 and an interface 20. The common time source 18 is preferably an independent clock, e.g., a clock synchronized with the atomic clock of the National Institute of Standards and Technology (NIST). The use of an independent clock allows maximum flexibility of the system. (Alternatively, as described later, the recording system 10 can use the clock of the POS system that is generated by the host computer 14.

The interface 20 receives the synchronizing signal from the common time source 18 and converts it to a form that can be recorded with the video signals generated by camera 16. The interface 20 may also add some static data to be recorded with the video signal to describe the store location, camera number/position, cashier's name/ID or other indicia.

Recording device 22 records the video signal from camera 16 and the coded data signals (including the synchronizing

signal) from the interface 20. The recording device 22 may be a video cassette recorder, video disc, or a computer-based video capture system. The present system may have more than one recording device 22. Typically, one recorder 22 is required for each camera. However, some camera systems use video combining and multiplexing techniques to allow a single recording device 22 to record information from multiple cameras.

The recording device 22 records the information from the interface in such a manner as to preserve the entire video signal generated by camera 16. For instance, if the recording device 22 is a video cassette recorder, the video signals are stored on the video portion of the video tape, while the coded data signals from interface 20 may be recorded on the audio portion or frame intervals of the video tape.

Regardless of how the synchronizing signal is recorded by the recording device 22, it should be noted that the synchronizing signal is simultaneously or synchronously recorded on the video tape with the video signal. That is, the synchronizing signal is used as a permanent indexer or marker to indicate the exact storage location of the video signals. Therefore, if the cameras sense a cashier ringing up the sale of a loaf of bread at 11:29 am, the video signal corresponding to the cashier's behavioral action is recorded on the video tape along with the synchronizing signal representing 11:29 am. During playback, every time the

synchronizing signal for 11:29 am is accessed, the video tape will display the cashier ringing up the sale of the loaf of bread. Similarly, the synchronizing signal is simultaneously stored with the transaction signal on the transaction database. During playback, every time the synchronizing signal for 11:29 am is accessed, the transaction database will call up the transaction signal corresponding to the description of the loaf of bread and the price. In this manner, the AVETDM system is able to synchronize the transaction events with the appropriate behavioral events during playback.

An alternative recording system 11 of the AVETDM system is shown in Fig. 2, in which like devices are similarly numbered. In Fig. 2, the point-of-sale terminal 13 generates its own synchronizing signal which is input into the interface 20 and the transaction database 14. In this embodiment, the synchronizing signal is a timing signal from an internal clock. Accordingly, an independent synchronizing signal generator is not required. However, other sequencing information generated by the POS terminal 13 may also be used (e.g., a transaction sequence number). Again, the POS generated synchronizing signal is used to synchronize the video or behavior information stored on recording device 22 with the transaction information stored on transaction database 14 during playback.

Referring now to Fig. 3, a second alternative recording system 15 is shown. Again, like elements are similarly numbered. In this embodiment, the interface 21 combines the video signal from camera 16 and the synchronizing data signal generated by the POS terminal 13 into a combined video/data signal. The recording device 22 records the combined video/data signal. It is preferable that the interface 21 combines the video and data information in such a manner that all of the information is preserved. If the recording device 22 is a VCR, the combined video/synchronizing signal is stored on the video portion of the video tape. The transaction database 14 stores all of the transaction data simultaneously with the synchronizing signal from POS terminal 13 in the manner described in the previous embodiments.

Referring now to Fig. 4, an embodiment is disclosed which allows twenty-four hour coverage of a plurality of POS stations for an extended period of time. Input lines 100, 102, 104... which carry the video signals from each camera are connected to a bank of recording devices 26 via splitters or T's 28. The splitters 28 divide the video signal from one camera to the recording devices. Timers associated with each recording device are programmed to turn on and off at appropriate times. In this situation, the recording device of choice is a VCR. Each VCR 26 can record eight hours of behavioral events occurring at a POS terminal. Therefore, a bank of three VCRs can provide twenty-four hour coverage for one camera.

The output line 106 from interface 20 transmits the synchronizing signal from the common time source 18. Splitters or T's 28 are used to input the synchronizing signal into an input channel of each VCR 26 from line 106. In the preferred embodiment, the input channel is an audio input of the VCR's. The embodiment disclosed in Fig. 4 can be used with the recording systems shown in Figs. 1 or 2.

In Fig. 5, a schematic diagram of an alternative recording system that utilizes a plurality of interfaces is shown. The recording system of Fig. 5 may also be used with the recording systems of Figs. 1 or 2.

The preferred configuration of the interface 20 is shown in Fig. 6. A configuration terminal 32 or other controller means is connected to a microprocessor 30. The configuration terminal 32 allows an operator to check the status of the interface or to select certain parameters. The synchronizing data, e.g. timing information from a clock, enters the microprocessor 30 at input 40. The outputs 107, 109 and 111 of the microprocessor 30 are connected to various data encoders 34, 35...M which convert the synchronizing signal into a format compatible for recording on the recording means 22. The type of data encoders 34, 35 used depends on the recorder 22. For example, if the recorder 22 is a VCR, the data encoder may be a Data Based Security, Moorestown, N.J., model AM 90.

Fig. 7 shows an alternate interface circuit 21, which is used in connection with the recording system embodiment as shown in Fig. 3. Again, the configuration terminal or other controlling system 32 is connected to the microprocessor 30 via line 42. The raw data, including the synchronizing data, coming from the POS terminal 13 is input at 40 to the microprocessor 30. The outputs 107, 109 and 111 of the microprocessors 30 are connected to the encoders 44, 45....N. The video signal from the camera 16 is also input into the encoder at 50. The inputs may all be from one camera (in which case 50a, 50b and 50n are connected to the lone camera by splitters or T's) or alternatively each video input 50a, 50b...50n may be connected to its own camera 16a, 16b...16n. The encoders 44, 45...N produce an encoded combination signal of the video and data which is compatible for storing on the recording means 22. Again, this encoded information is generated in such a way that the entire data information and video information are not degraded. The outputs 114, 116 and 118 of the encoders are connected to the input of the recording device 22.

Fig. 8 shows an alternative embodiment which will allow an extended time coverage of the POS terminals for use with the system shown in Figs. 3 and 7. In this case, the behavior data generated by camera 16 is combined with the transaction data before recording on VCR 23. Splitters 28 direct the combined signal from the interface to each bank of VCRs.

Referring now to Fig. 9, an independent time source, as may be used in the system of Fig. 1, is shown. In this example, the time source 61 is the atomic clock supported by the National Institute of Standards and Technology. The time source 61 generates a real time signal. The time source 61 is connected to a transmitting antenna 62, which transmits, via wireless technology, the time signal to receive antennas 64. Receiving antennas 64 amplify the received timing signal and forward the timing signal to a time receiving box 66. The time receiving boxes 66 condition the timing signal in manner that the POS terminal 12 and the interface 20 can better utilize the timing signal. A serial interface (not shown) may be used to connect the time receiving boxes to the POS terminal 12 and the interface 20. In this manner, the transaction database 14 and the recording mechanism 22 store the exact time simultaneously. Upon playback of the transaction data and the recording medium, the timing signal will be used to synchronize the transaction history with the behavioral history.

In Fig. 10, an alternative common time source is shown. In this system, a master time receiver 70 is connected to receive antenna 64. The output 71 of the master time receiver 70 is connected to a plurality of slave time receivers 72. The slave time receivers 72 are connected to the POS terminal 12 and the interface 20. Again, the transaction database 14 and the recording mechanism 22 simultaneously record the exact time with the transaction signals and the video signals,

respectively, which is used to synchronize the data during playback.

Fig. 11 is a second alternative common time source circuit. The time source 61 is connected to a modem 74. The modem 74 is connected in the usual manner via telephone lines to a public telephone network 76. The surveillance system also has modems 78 and 79 which are connected to the public telephone network 76. The modems 78 and 79 are connected to the POS station 12 and the interface 20 respectively.

The AVETDM playback system is shown in Fig. 12. The transaction database 14 is connected to a compatible transaction database 80. Compatible transaction database 80 converts the transaction information and the synchronizing signal stored on the transaction database 14 during the recording period into a compatible format. The compatible transaction database 80 is connected to a microprocessor 82. The microprocessor controller 84 is directly connected to the microprocessor 82. In the preferred embodiment, the compatible transaction database 80, controller 84 and the microprocessor 82 are an off-the-shelf computer, for example, an IBM compatible 486 computer. In this case, the controller 84 would include the terminal and/or keyboard of the computer, and the transaction database 14 is a floppy disk, magnetic tape or other media which can be read by the computer.

The transaction signal and the synchronizing signal from the compatible transaction database 80 is downloaded into a memory or database of the microprocessor 82. This information is then accessed by the microprocessor 82 at the appropriate time. An alternative system (not shown) can have the microprocessor 82 controlling the transaction database 14 to download the data when needed.

An output of the microprocessor 82 is connected to a playback device 86. The playback device 86 must be compatible to the format used by the recording means 22 to record the video signal generated by the camera. For example, if the recording means 22 is a VCR, the playback means 86 is preferably a VCR also. An output of the playback mechanism 86 which can access the stored synchronizing signal (e.g., an audio output) is connected to a decoder 88. The decoder 88 decodes the synchronizing data from the playback device 86. An output of the decoder 88 is then connected to the microprocessor 82.

The stored video signal from camera 16 is reconstituted by the playback device 86 and input to the overlay control box 90 via line 142. The video signal contains the behavioral history which occurred at the POS station 12 (e.g., a person purchasing their groceries or a truck driver paying a toll). The microprocessor 82, using the synchronizing signal played back from both sources (i.e., the synchroning signal which was stored in the transaction database 14 and the synchroning

signal stored by the recording device 22) synchronizes the behavioral/video information with the appropriate transaction data. The microprocessor 82 outputs a video overlay signal to an overlay control box 90 via line 140. The video overlay signal is an alphanumeric display of the transaction data corresponding to the exact behavioral events being played by playback means 86. The overlay control box 90 produces a composite video signal which includes the behavioral history being overlaid with an alphanumeric display of the corresponding transaction data. This composite signal is then displayed on monitor 92.

An operator at controller 84 decides which behavioral events stored by the recording device 22 and which transaction events stored on the transactional database 14 to view. To promote easy operation, the controller 84 may be menu driven.

An example of the AVETDM system's operation will now be given. Referring now to Fig. 13, a surveillance system according to the present invention is shown that is designed for use with multiple cameras and multiple point-of-sale terminals, and is compatible with the system of Fig. 1. Each camera 126, 127, 128, 129,... is positioned to cover one cash register or point-of-sale terminal 112, 113, 114, 115,..., respectively.

The cameras sense the video images (behavioral events) as they transpire at each POS terminal. The images are converted

into a video signal capable of being recorded by recorders 134, 135, 136. Each recorder simultaneously records the synchronizing signal from the common time source 130.

(In the preferred embodiment, the recorders 134, 135, 136...also record the transaction history from the POS host computer 116 along with any identifying indicia and the synchronizing signal. Although this step is not necessary, it is useful to have verifiable evidence for any future judicial proceedings. The transaction data is stored so as not to interfere with the recorded video signal. If the recorder is a VCR, the recording medium is a video tape. The transaction data can, for example, be stored on the video frame intervals or on the audio channel portion of the video tape. Therefore, the entire behavioral history and transaction history is recorded and may be accessed at a future date.)

All POS terminals are connected to a POS host computer 116. The POS terminals 112, 113, 114, 115... send transaction data to a POS host computer 116 in bursts when the final total is determined for each customer. For this example, the transaction data corresponding to each individual item or good rung up by the cash register is stored (e.g., in a buffer or in RAM) by each POS terminal until all items for that particular customer have been rung up. That is, after the cashier pushes the tender or total key on the cash register, the POS terminal sends a transaction signal corresponding to the transaction history to the POS host computer 116.

The POS host computer generates a database for each POS terminal containing the transaction data, any identifying indicia and the synchronizing signal from the common time source 130. The POS host computer 116 stores this database on a permanent media (magnetic tape, CD ROM, floppy disks, etc.). The information stored on the floppy disks forms the transaction database 14 of Fig. 1.

The interface 132 operates in a similar manner as the interface described in Figs. 1 and 6. Similarly, the recorders 134, 135, 136... also operate in a similar manner as the recorder 22 described in Fig. 1.

During playback, the operator may desire to view the events at a particular POS terminal, e.g. POS terminal 113. The operator inputs this information into the controller 84. The microprocessor 82 controls the playback device 86 to output the video signals corresponding to POS terminal 113. Normally, all of the transaction data for all POS terminals (including the synchronizing signal and identifying indicia) is downloaded into a buffer of the microprocessor 82 from the transaction database 14 or floppy disk. By matching the synchronizing signals from the playback device 86 with the synchronizing signals stored in the buffer, the microprocessor recalls the exact transaction signal which corresponds to the behavior events at POS terminal 113 at the precise point in time at which the video signals were stored on the video tape. The microprocessor then generates a video outlay signal

corresponding to an alphanumeric display of the desired transaction data of POS terminal 113. The overlay control box 90 then overlays the video outlay signal over the corresponding video signal from playback device 86, generating a composite video signal. The operator sees the behavioral events of POS terminal 113 and the corresponding transaction history overlayed on the video.

The microprocessor has both synchronizing signals, which were stored during the recording session (see Fig. 1), available to process. That is the synchronizing data stored by the transaction database 14, and the synchronizing data stored on the video tape by the recording mechanism 22 are both input to the microprocessor 82. Therefore, during playback the microprocessor 82 is able to match or synchronize the behavioral events (video information) with the later stored transaction events.

Since the microprocessor 82 has available to it <u>all</u> of the transaction data recorded onto the transaction database 14, it is possible for the operator to look into the "future". For example, the operator may wish to look at all transactions at POS terminal 113 in which a customer uses a credit card. The operator can direct the microprocessor 82 to look for identifying indicia corresponding to credit card sales. Therefore, while the playback device 86 is showing behavioral events taking place at the "present" time, the microprocessor 82 can inform the operator that a credit card transaction will

take place in twelve minutes and thirty-one seconds (or alternatively, the eighth customer from now will use a credit card).

It should be noted that the transaction data may also be stored with the video images. The surveillance system allows the operator to freeze the alphanumeric transaction data display in the overlay while the video tape is rewound to the beginning of the transaction. The operator can then review the behavioral events while looking at the transaction data in the overlay.

Line markers are provided on the overlay to enable the user to move a pointer on the overlay to each item as it is registered by the cashier in the picture. The system encodes the next transaction serial number on the sound track of the video tape at the end of each transaction. Alternatively, an independent clock can be used to synchronize the data stored on video tape with the data stored in the second medium. The user can use the serial number to be certain that the transaction behavior that he is seeing corresponds to the transaction data in the data overlay.

Another common use of the AVETDM surveillance system is at toll booths. In certain toll systems, the toll transaction data is stored in a lane controller. The transaction signal may be generated by a toll terminal, card reader, loop detector, treadle, indicator light, etc. located at each toll

lane. After the data is stored in the lane controller, it is sent to the plaza controller at irregular intervals. Several transactions take place between data transmissions. For example, five vehicles may have to exit the lane before a transaction is transmitted to the host computer. If several transactions occur before the data is transmitted, there is no method of synchronizing the transaction data with the video images other then a data freeze or data pause method. However, the data pause method is not acceptable to toll authorities, since it requires too much time to review tapes and does not work well with automated tape editing. The solution is to develop circuitry which inputs synchronizing data, encodes it, and stores it along with the video signal. The synchronizing data may be stored on the video tape's sound track, the video tape's vertical interval or within the visible video. The transaction data is stored asynchronously from the behavioral events. However, the transaction data is also stored with the synchronizing signal in a manner that supports synchronizing or resynchronizing the recorded behavior with the transaction data upon playback.

The AVETDM computer encodes the video tape with identifying indicia, including facility name/number, camera identification, date and time messages. The time messages can be inserted every 100 milliseconds. The AVETDM simultaneously captures the transaction data from the lane controller. These messages include several transactions. Each transaction includes time messages for that transaction. Each transaction

message can be stored in a file with the date, lane, terminal identification, etc., and the time.

The file will contain a list of messages indexed by date, time, lane, etc. Upon replay,the user selects the target lane. When the video tape is replayed, the AVETDM inputs the data and time messages. The AVETDM requires the time messages from the tape and looks up the messages in a data file on the disk for that time and target lane. The AVETDM then displays the messages from the targeted lane and the data overlay. In one embodiment, the lane controller stores three transactions and forwards all three at the end of the third transaction to a host computer.

A camera is placed at the toll plaza to view four to six lanes from the exit side. A video tape recorder, for example a VHS video cassette recorder, is connected to each camera. The AVETDM computer is connected to the lane controller communication so that the AVETDM computer receives all of the data messages that are communicated between the lane controller and the plaza computer. The VHS tape is encoded with data messages that identify the plaza and the date. Time messages are encoded every 100 milliseconds. The plaza number, lane number, date and time are stored to the host computer's disk for every transaction message.

When the video tape is replayed, the operator indicates which lane is to be targeted for data to be displayed in the

overlay. When a tape is replayed, the time, date, plaza i.d., etc. are recorded on the sound track of the video and are read into the microprocessor 82. The microprocessor 82 reads the time and searches the data file for messages that have the targeted lane identifier and the matching time and date. When the messages are located, they are formatted and displayed in the data overlay. The time of the behavior in the picture now matches the display in the overlay.

Another situation where the AVETDM is necessary is where the POS system data is held until the end of the day or until the POS is closed. Either way, the transaction data is not available in real time. Consider a POS system that holds each transaction inside the POS terminal, and a host collects the held data only after the POS terminal is closed. The data is then stored in the POS system main processor and is made available to the AVETDM via a set of magnetic disks. Time synchronization between the POS system and the AVETDM is achieved by a direct serial link between the POS system main processor and the AVETDM.

The AVETDM will record the video images (behavioral events) of the POS terminal activities but will not be able to record any real time transaction data, as it has not yet left the temporary storage in the POS terminal. Instead the AVETDM records time and identity codes (synchronizing indicia) on the video tape which can be retrieved during playback. These codes will be used to synchronize the transaction and event

data received from the POS system with the video images of the POS terminal. Messages coded on the video tape can take the following form:

TIM yyyymmdd hhmmss zz<EOM>

IDN aaaaaa bbbbbb cccccc zz<EOM>

where,

TIM is the message header for the time message

IDN is the message header for the identity message

yyyymmdd is the date, with yyyy being the year, mm the month, and dd the day

hhmmss is the time, with hh being the hours, mm the minutes, and ss the seconds

zz is the message checksum

<EOM> is the end of message character, which in this case is a line feed (OxOB)

aaaaaa is the customer identity code

bbbbbb is the site identity code

cccccc is an identity code used to identify different systems within a site

The serial interface between the POS system main processor and the AVETDM uses the same time message format that is encoded on the tape. This message originates from the POS system main processor on a frequent periodic basis. For example, many systems make the time and date available when the receipt is printed. The AVETDM receives the message,

verifies the checksum, and makes any necessary adjustments to its internal clock. The AVETDM internal clock will thus be synchronized with the POS system clock. The AVETDM will use its internal clock to generate the time messages that are encoded on the tape.

The transaction and event data files consist of individual records detailing each transaction or event. Each of these records must contain a time stamp to allow them to be synchronized with the video stamp and therefore to be synchronized with the video tape. The playback system reads the data file before the tape begins to play. It extracts time data and other identifiers from the message and compares these with the information initially received from the tape as it begins to play. This information will verify that the tape is in fact the one recorded for this particular data. As the tape plays the operator is given details on each transaction at the appropriate time.

The entire transaction data can be loaded into memory before the behavioral recording is replayed. The stored transaction data can then be manipulated by the computer to provide summaries, averages, counts, statistics, anomalies or exceptions. The desired information or anomalies can then be inserted during the replay of the behavioral transaction at the relevant times or sequences. The system can "look into the future" and display a warning message to the operator that a particular anomaly or transaction will occur in X number of

frames or in X number of minutes. A control program can be customized to each user to look for particular events and to support this level of operation, thereby allowing the operator to see transaction data derived from past, present and future behavioral events while viewing the "present" provided by the video tape replay.

An added benefit of the AVETDM system is that because the data is known in advance, the system can warn the operator before items of interest appear on the tape. The general format of the transaction and event data records is as follows:

TRN yyyymmdd hhmmss tttttttttt....tttt zz<NL>

The specific details of the transaction record such as type of sale, amount, specific details about the items purchased, etc... are not important to the description of the AVETDM system and are simply represented as a series of t's above. The AVETDM will process these specific details in order to extract the information that is to be displayed to the AVETDM system operator.

The transaction record must contain a means for synchronizing the transaction data with the behavioral events recorded on the video tape. The preferred method is to time stamp the transaction data and the recorded behavioral events which allows video and data synchronization. However, other

methods of synchronizing may be used. For example, pictorial or graphic recordings may be used. The transaction data system buffers some number of transactions before transmitting to a host computer. If four transactions are buffered, the first transaction is transmitted from the local buffer when the fifth transaction is consummated, the second transaction is sent after the sixth transaction is consummated and so forth. In these instances, the AVETDM can receive the signal that a transaction is being sent and can mark the tape with sequence codes. When the data for the first transaction is sent, the code written on the behavioral recording matches the behavior that coincides with the fifth transaction and so on.

Additional identification details (such as store and customer) might be contained in each transaction record or might appear in a special file identity and description block at the beginning of the file. Either way the correlation between the file and tape can easily be verified.

Another typical situation where the AVETDM is necessary is one where the POS system data is not available in real time, but is available at a slightly later time.

Consider a POS system which does not have its own internal clock or does not produce appropriate synchronizing indicia. Again, this example will use a POS system which holds each transaction inside the POS terminal, and collects the held data only after the POS terminal is closed, e.g., at

the end of the day. The data is then stored in the POS system host computer and is made available to the AVETDM via a set of magnetic disks. Time synchronization between the POS system and the AVETDM is achieved via a time signal receiver. Both the POS system and the AVETDM will connect to time receivers that receive a broadcast time synchronization signal such as the one transmitted by the NIST on WWV and WWVH. (See, for example, Figs. 9, 10 and 11.)

The AVETDM will record the video images of the POS terminal activities but will not be able to record any real time event and transaction data, as it has not yet left the temporary storage in the POS terminal. Instead the AVETDM records time and identity codes on the video tape that can be retrieved from the POS system with the video images of the POS terminal. Messages coded on the video tape can take the following form:

TIM yyyymmdd hhmmss zz<EOM>

IDN aaaaaa bbbbbb cccccc zz<EOM>

where,

TIM is the message header for the time message

IDN is the message header for the identity message

yyyymmdd is the date, with yyyy being the year, mm the month, and dd the day

hhmmss is the time, with hh being the hours, mm the minutes, and ss the seconds

zz is the message checksum

<EOM> is the end of message character, which in this case is a line feed (OxOB)

aaaaaa is the customer identity code

bbbbbb is the site identity code

cccccc is an identity code used to identify different systems within a site

The serial interfaces with the time signal receiver use the same time message format that is encoded on the tape. This message originates from the time signal receiver on a frequent periodic basis. The AVETDM and POS system receive the message, verify the checksum, and make any necessary adjustments to their internal clocks. The AVETDM internal clock will synchronized with the POS system clock. The AVETDM will use its internal clock to generate the time messages that are encoded on the tape an the POS system will use its clock to time stamp all transaction and event records.

The transaction and event data files consist of individual records detailing each transaction or event. Each of these records must contain a time stamp to allow time to be synchronized with the video tape. The playback system reads the data file before the tape begins to play. It extracts time data and other identifiers from the message and compares these with the information initially received from the tape.

This information will verify that the tape is in fact the one recorded for this particular data. As the tape plays the operator is given details on each transaction at the appropriate time.

Even though particular embodiments of the present invention have been illustrated and described herein, this is not intended to limit the invention. It is therefore to be understood that modification and variation of the embodiments described above may be made without departing from the spirit or scope of the invention.

* * * * * * *

CLAIMS

I claim:

1. A surveillance system which asynchronously records digital data with respect to the video data, comprising:

   a) light sensing means for generating video signals of behavioral events corresponding to a transaction at an operation station;

   b) sensor means at the operation station for generating digital signals representing transaction events;

   c) means for generating synchronizing signals;

   d) first recording means for storing the video signals and the synchronizing signals;

   e) second recording means for storing the digital signals and the synchronizing signals;

   f) first playback means for retrieving the video signals and synchronizing signals stored on the first recording means;

   g) second playback means for retrieving the digital signals and synchronizing signals stored on the second recording means; using the synchronizing signals to synchronize the video signal with the digital signals;

   h) control means, responsive to an input signal, for generating a composite video signal, the composite video signal including signals representing alphanumeric displays corresponding to desired

transaction events, and alpha-numeric display is overlaid on the video signal of the desired behavioral events; wherein the control means utilizes the synchronizing signals to retrieve the desired behavioral events and transaction events; and

i) a monitor for displaying the composite video signal.

2. The surveillance system of claim 1 wherein the sensor means is a cash register.

3. The surveillance system of claim 1 wherein the means for generating synchronizing signals is a clock which can receive signals from an independent source.

4. The surveillance system of claim 1 wherein the first recording means is a video cassette recorder.

5. The surveillance system of claim 4 wherein the playback means is a video cassette recorder.

6. The surveillance system of claim 1 wherein the second recording means is a first computer and the information is stored on a magnetic medium.

7. The surveillance system of claim 6 wherein the second recording means is a second computer.

8. The surveillance system of claim 1 wherein the control means is a second computer.

9. The surveillance system of claim 1 wherein the first recording means is a video disc.

10. The surveillance system of claim 1 wherein the second recording means is a computer and the information is stored on a compact disc.

11. The surveillance system of claim 1 further comprising computer means associated with the control means for manipulating, calculating, sorting and filtering the stored digital transaction data for generating statistics and prompting for events that will be viewed on the replay behavioral record.

12. The surveillance system of claim 11 wherein the synchronizing signals are signals corresponding to the time elapsing when the recordings were made.

13. The surveillance system of claim 11 wherein the light sensing means is a television camera.

14. The surveillance system of claim 13 wherein the first recording means is a video cassette recorder.

15. The surveillance system of claim 11 wherein the synchronizing signals are sequence signals.

16. A method of asynchronously recording events including a video record and a transactional record, comprising the steps:

a) generating frames of video signals corresponding to the behavioral events at an operation station;

b) generating digital signals at the operation station for representing transaction events;

c) generating synchronizing signals;

d) storing the video signals and synchronizing signals on a first medium;

e) sorting the digital signals and the synchronizing signals or a second medium;

f) retrieving the stored video signals, the stored digital signals an the corresponding stored synchronizing signals;

g) synchronizing the video signals with the digital signals using the retrieved synchronizing signals;

h) generating a composite video signal, the composite video signal including signals representing alphanumeric displays corresponding to desired transaction events, said alphanumeric display is overlaid on the video signal of the desired behavioral events; and

i) displaying the composite video signal.

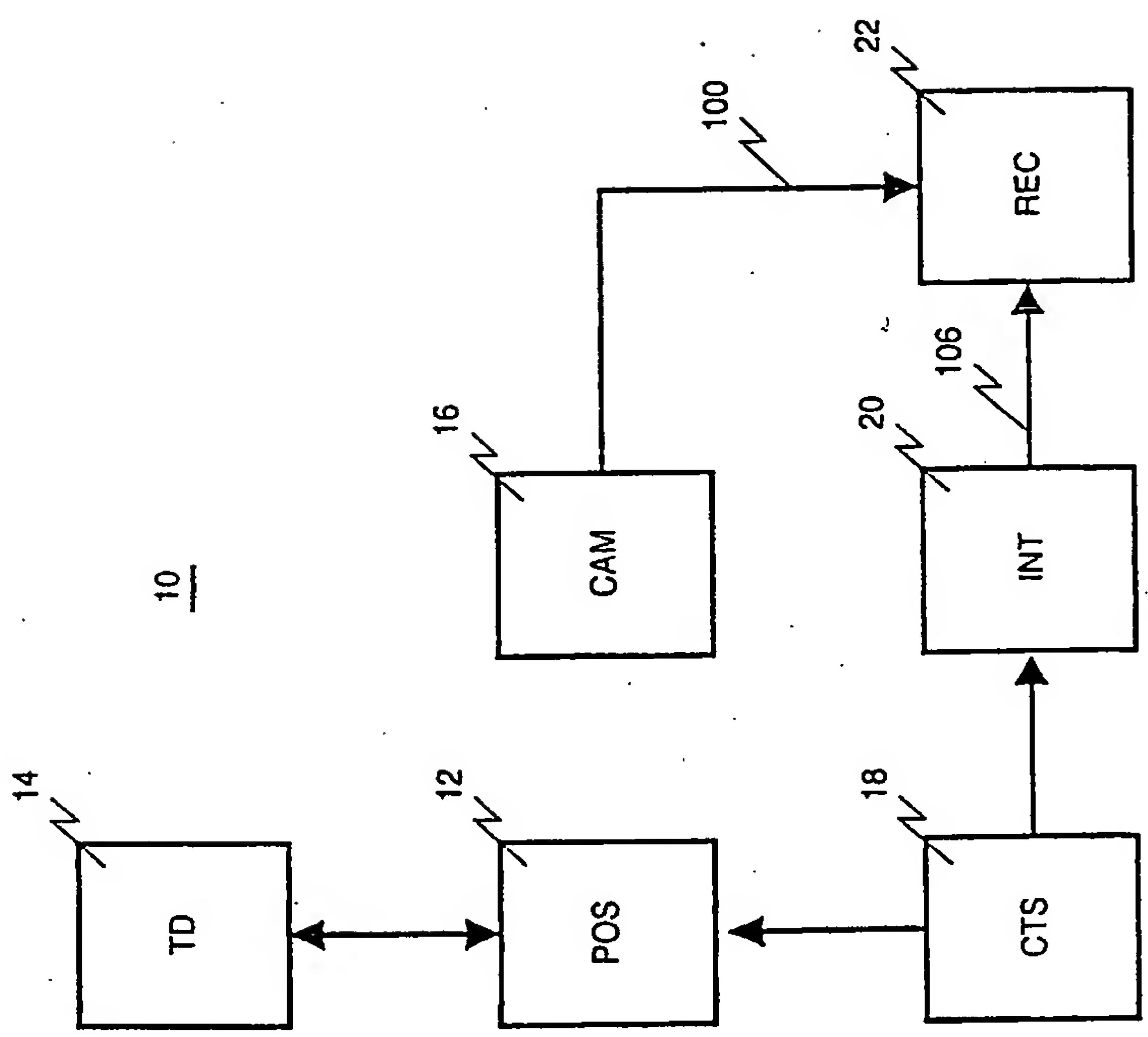


FIG. 1

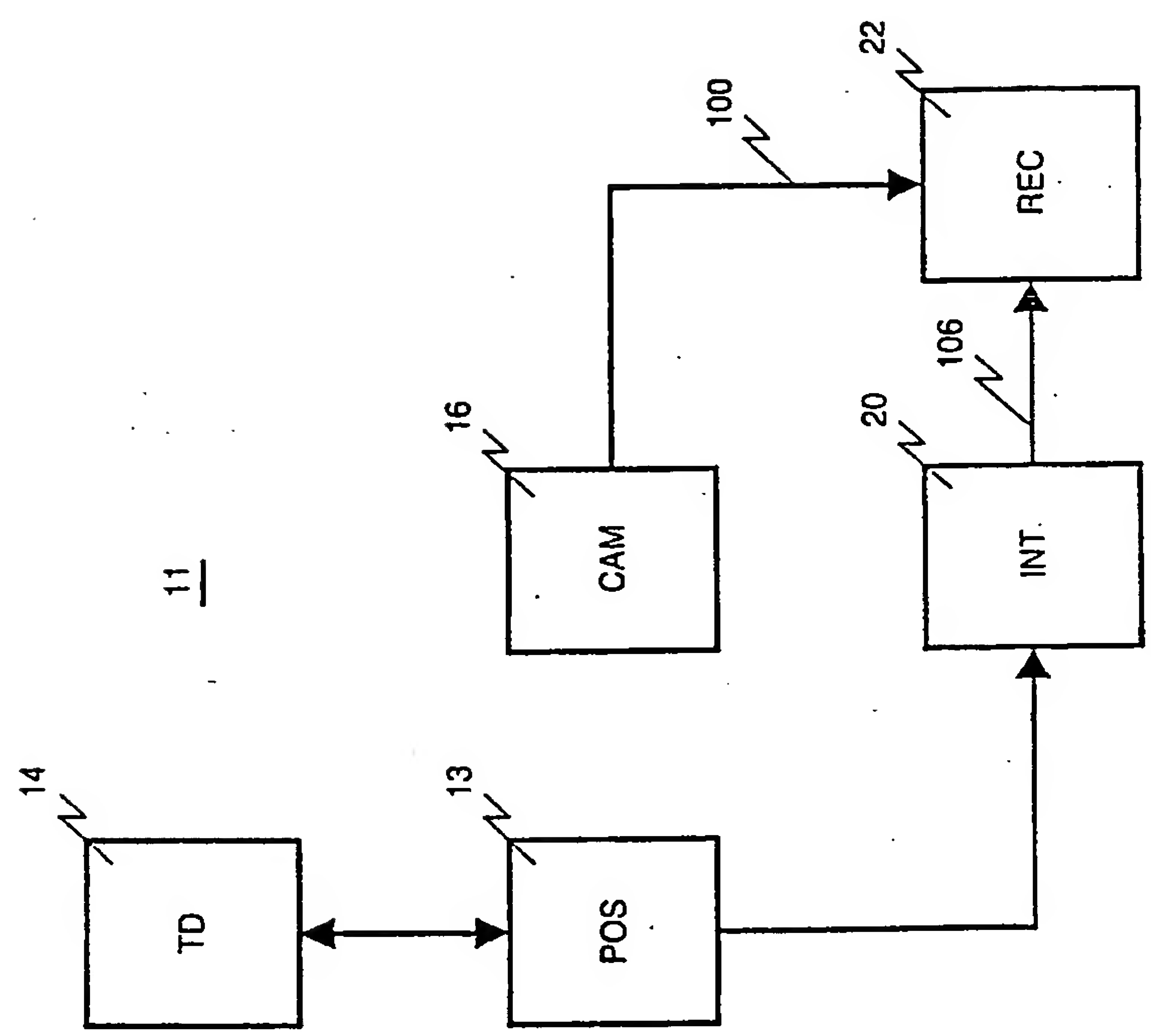


FIG. 2

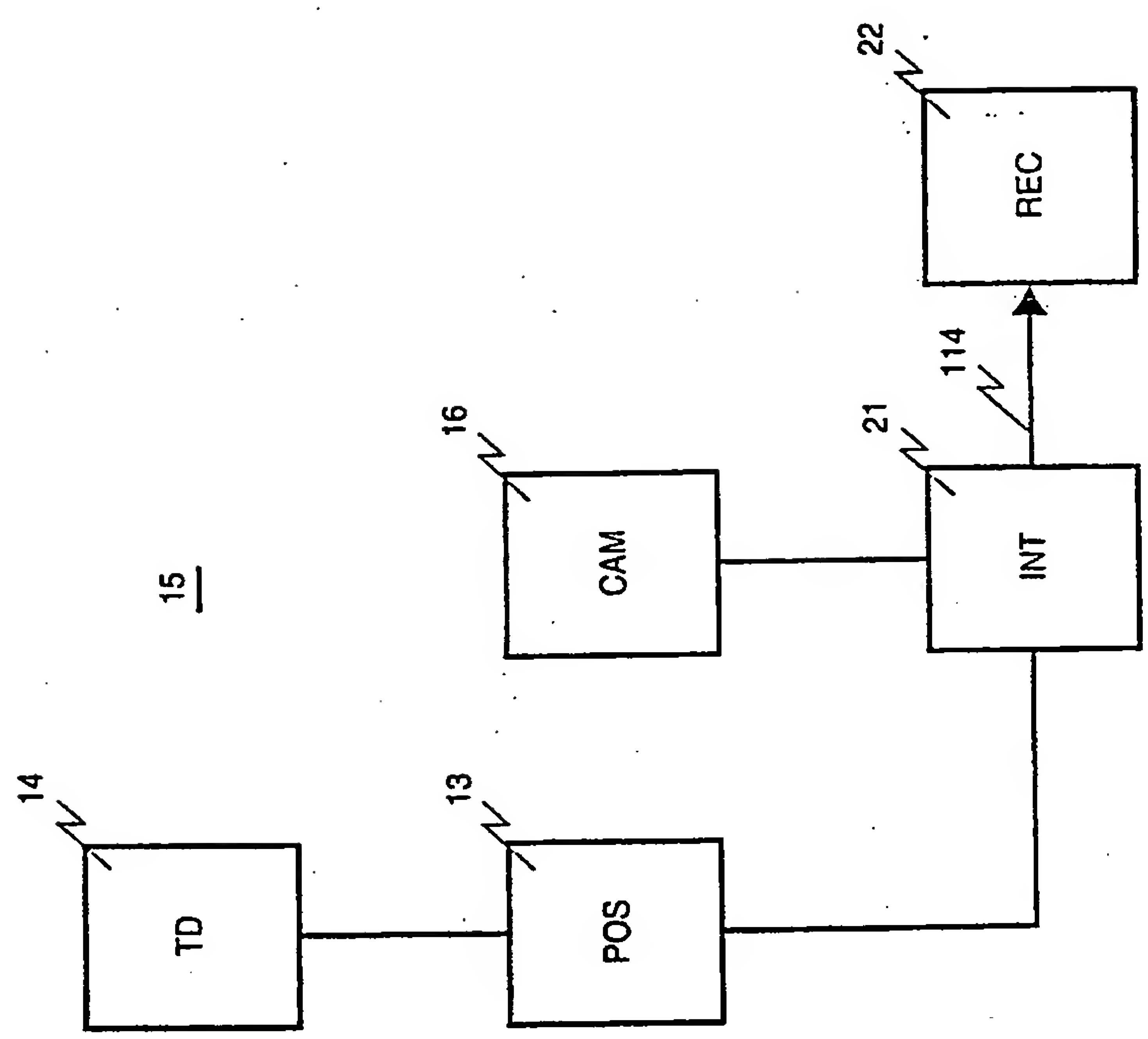


FIG. 3

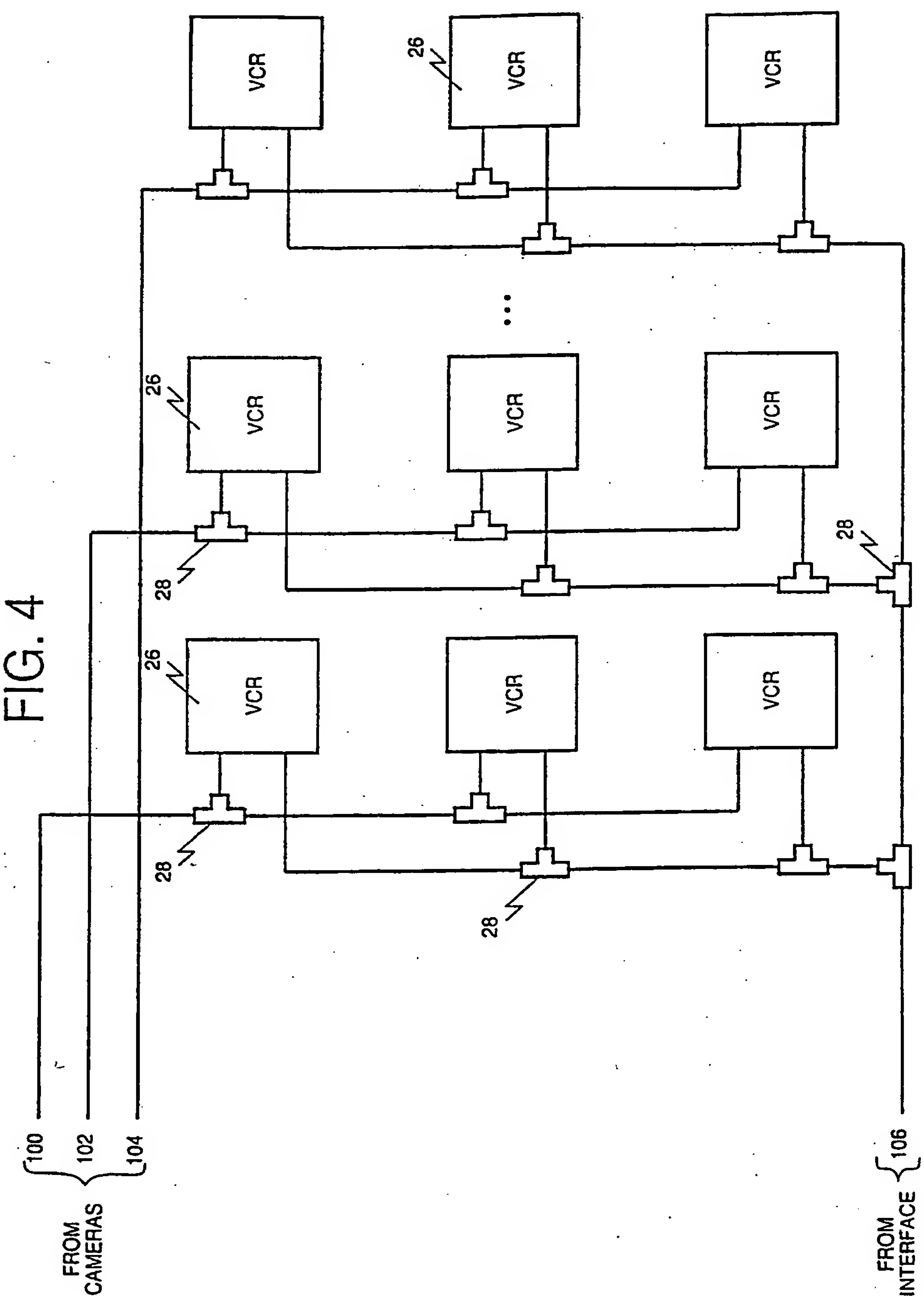


FIG. 4

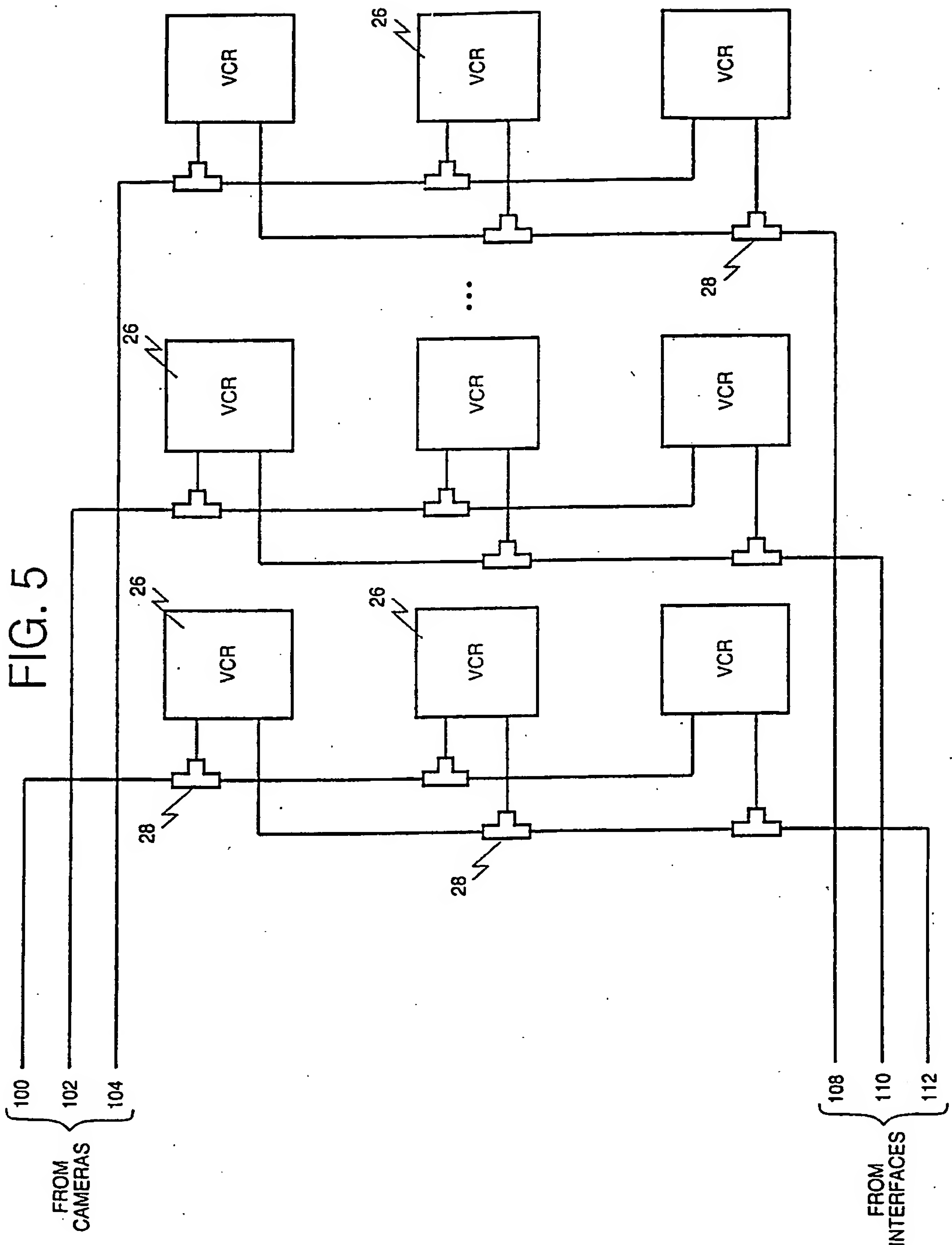

VCR
26
28
100
102
104
FROM CAMERAS
108
110
112
FROM INTERFACES

FIG. 5

FIG. 6
20

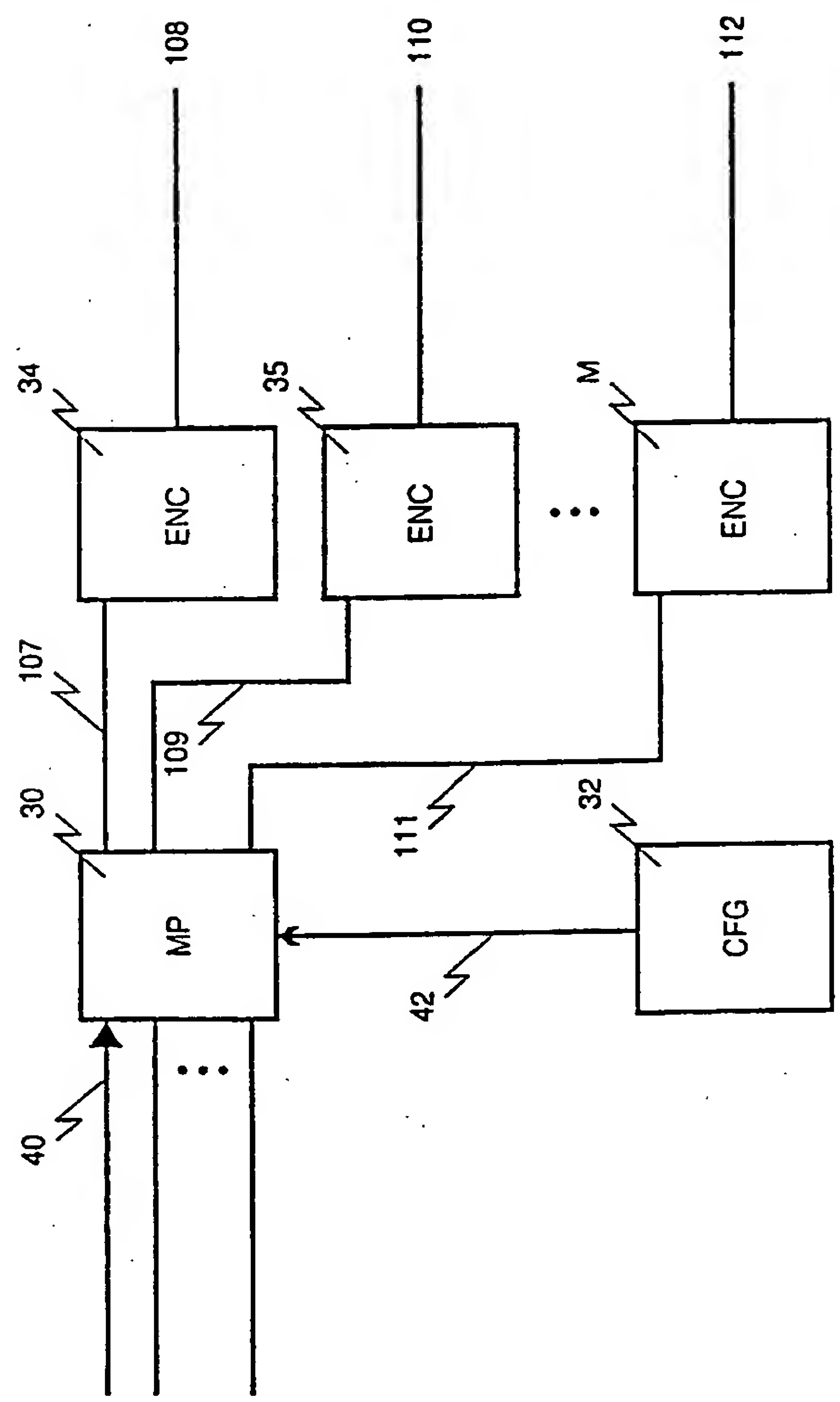
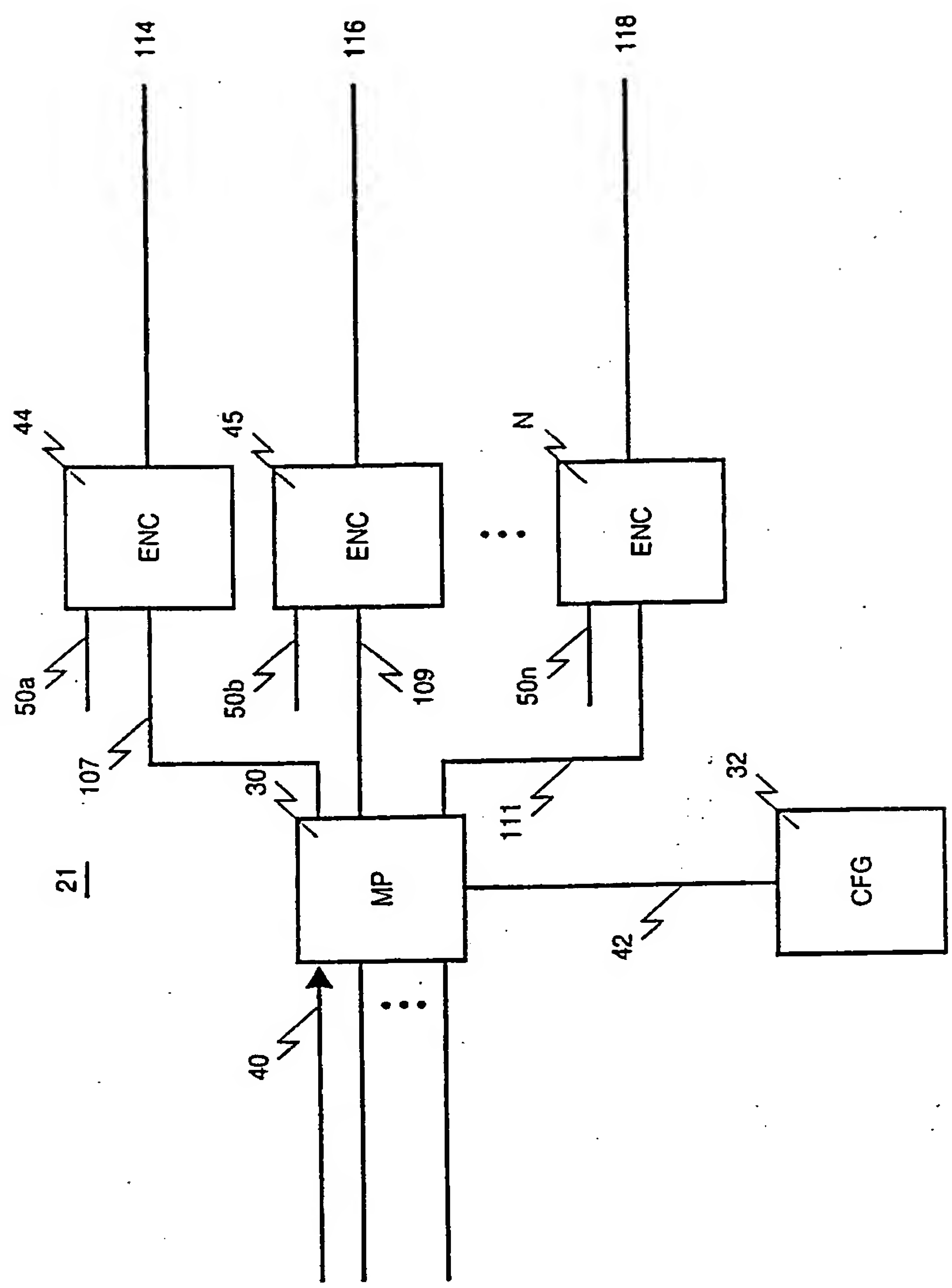


FIG. 7


114
116
118
44
45
N
ENC
ENC
ENC
50a
50b
109
50n
107
30
111
32
21
MP
CFG
42
40

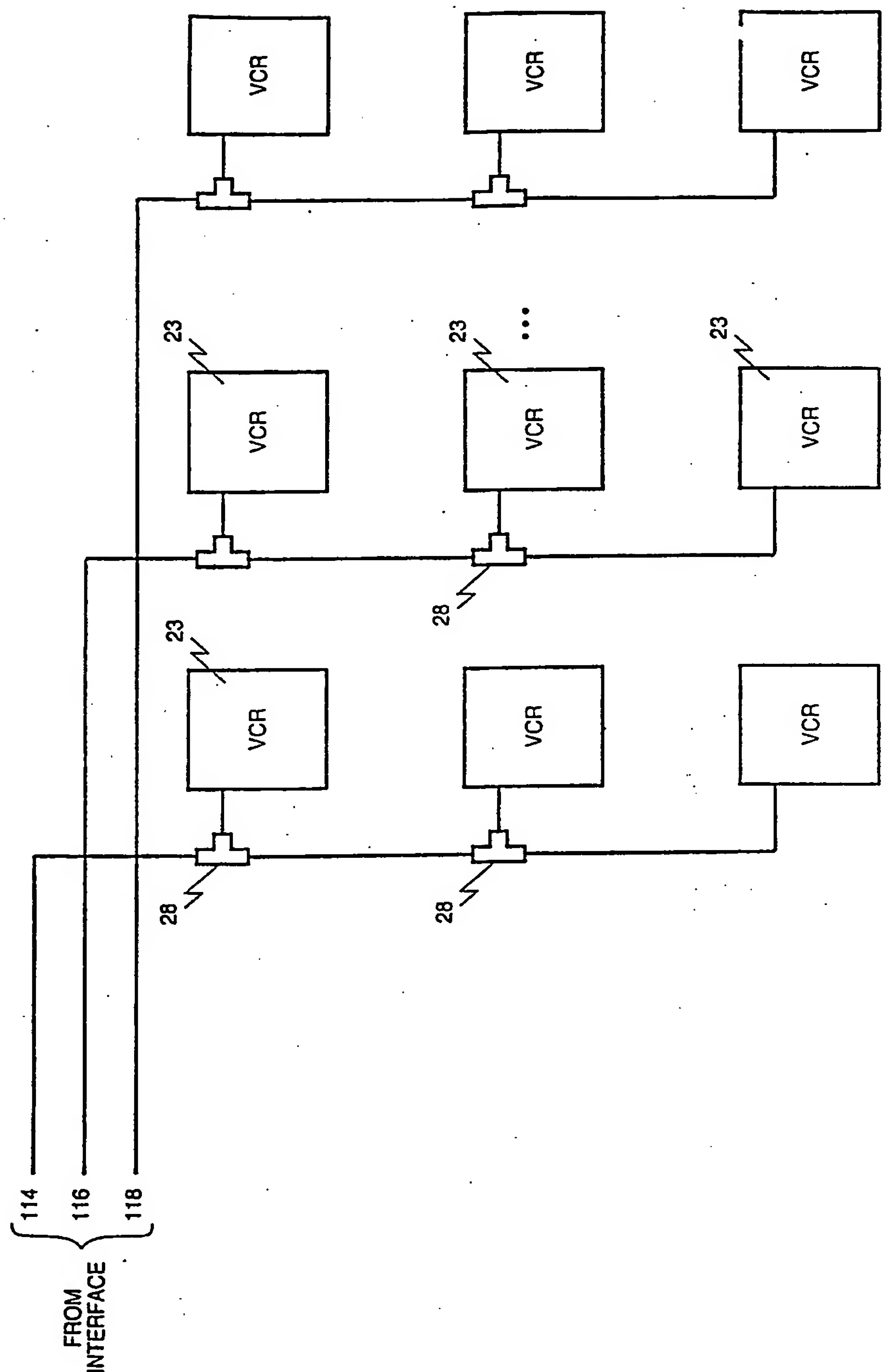

VCR
VCR
VCR
VCR
VCR
VCR
VCR
VCR
VCR
23
23
23
23
28
28
28
114
116
118
FROM INTERFACE

FIG. 8

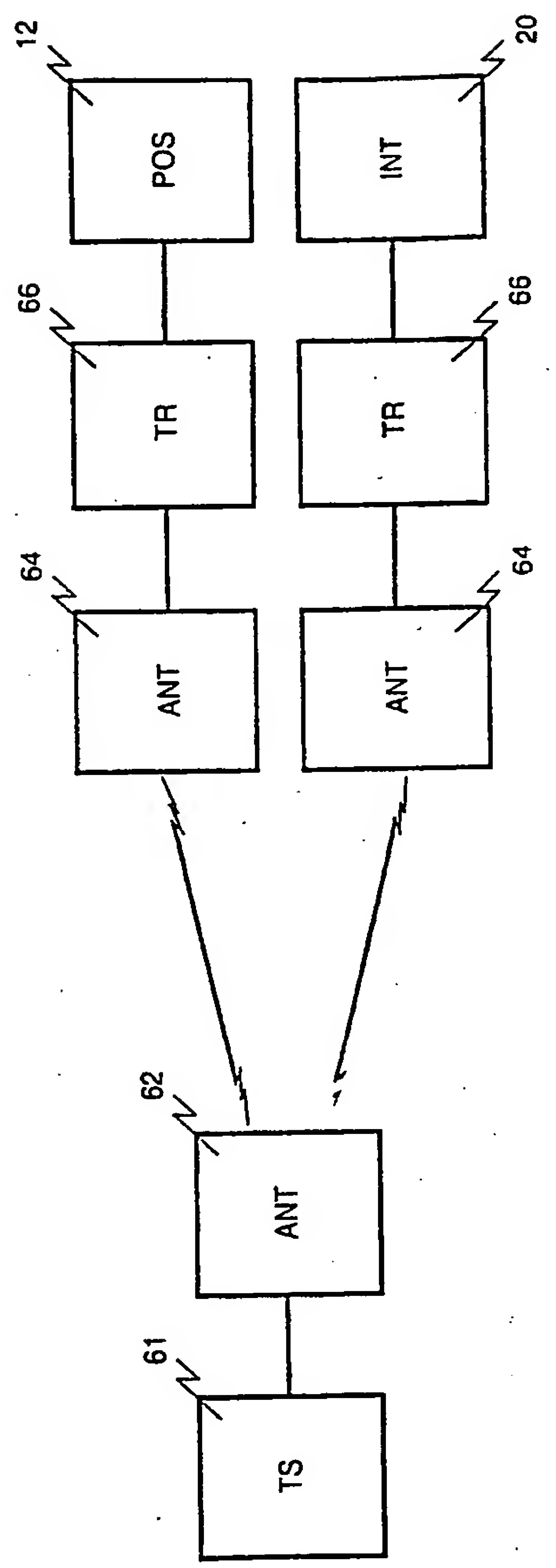


FIG. 9

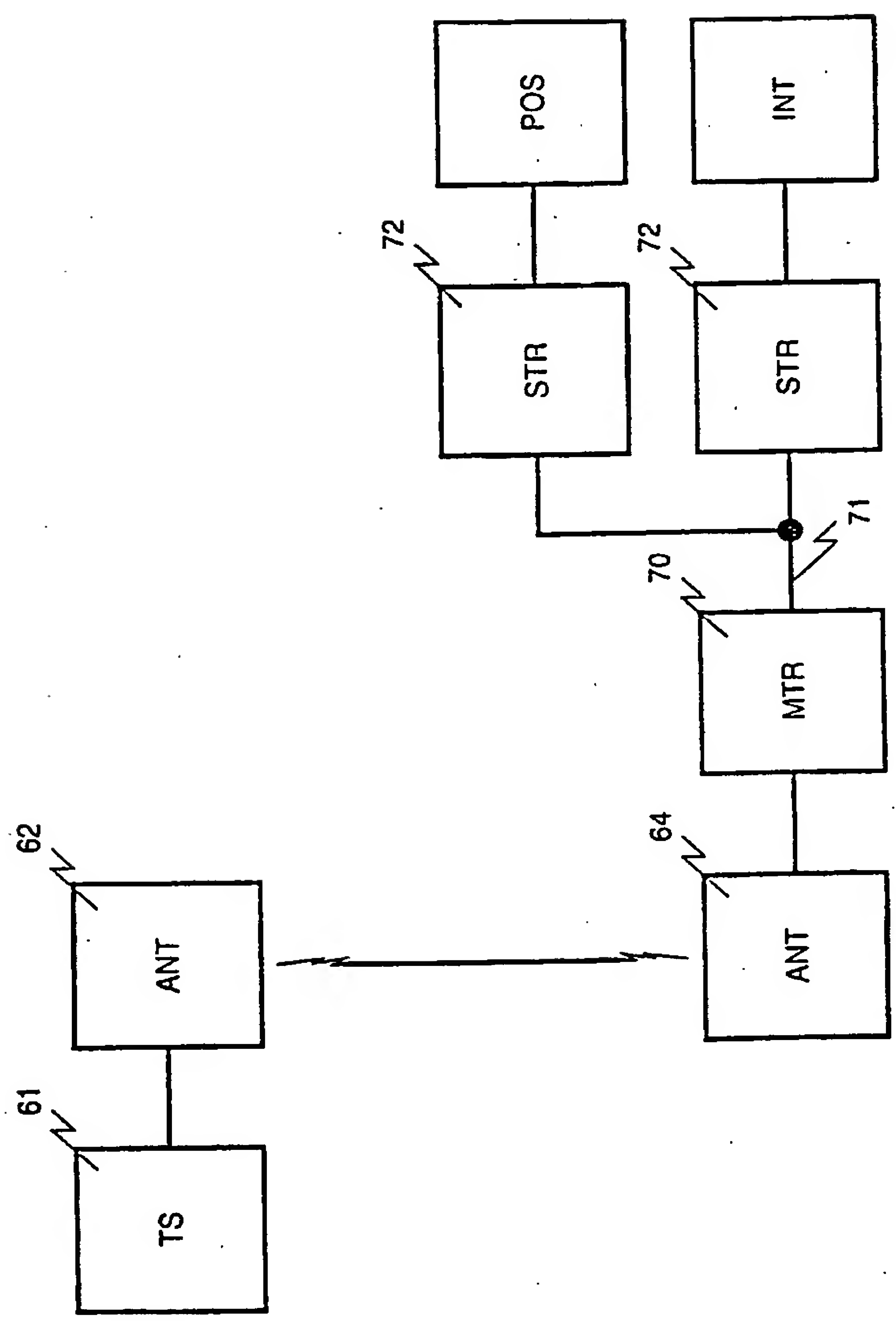


FIG. 10

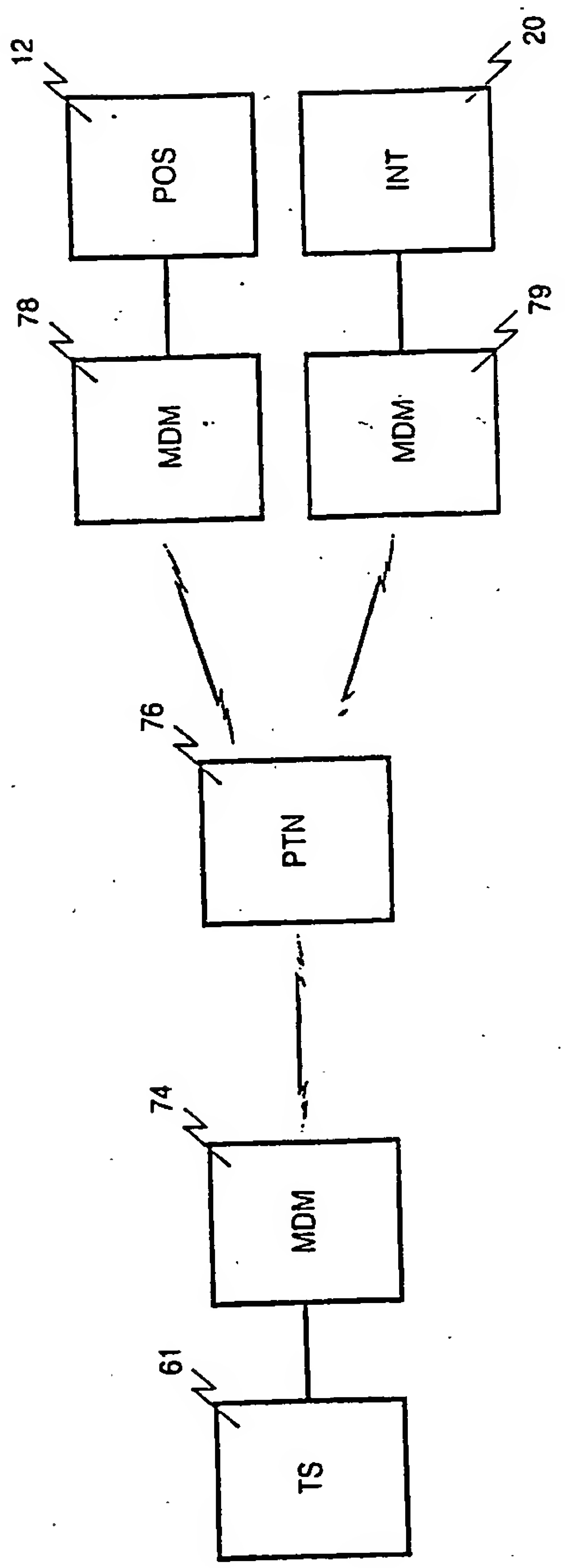
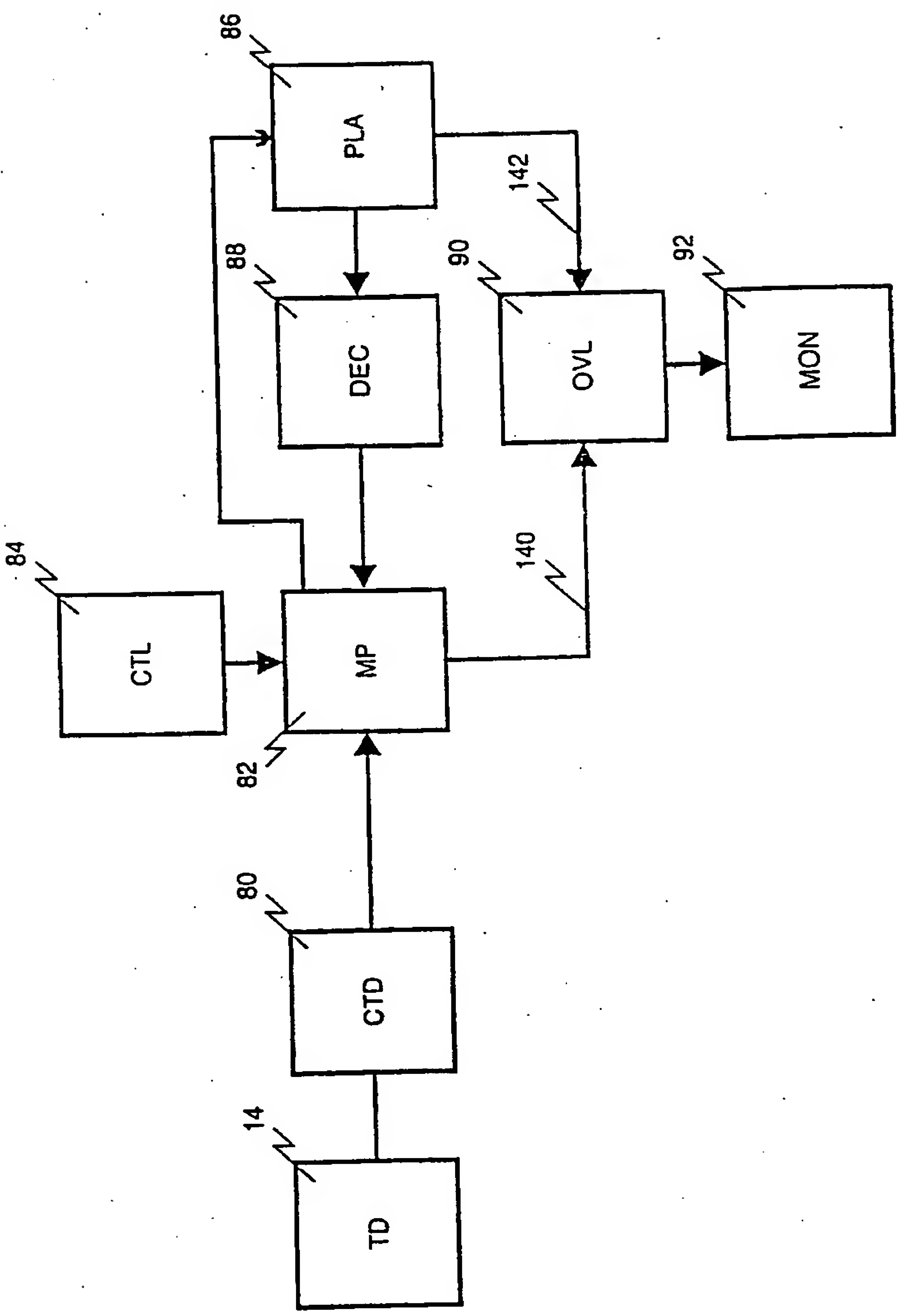


FIG. 11

FIG. 12



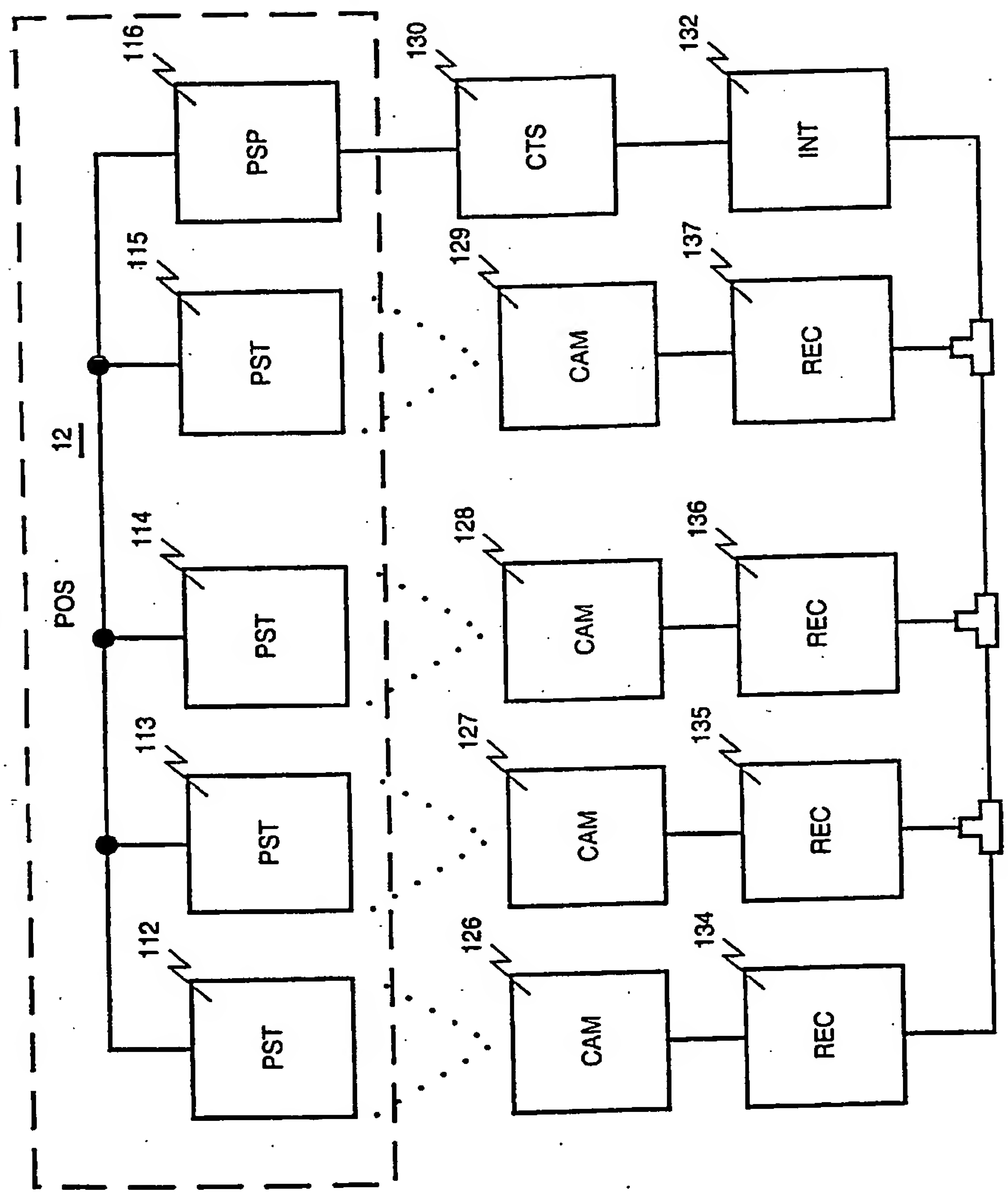

POS
12
116
PSP
115
PST
114
PST
113
PST
112
PST
130
CTS
129
CAM
128
CAM
127
CAM
126
CAM
132
INT
137
REC
136
REC
135
REC
134
REC

FIG. 13

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 95/05167

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G07G3/00 H04N7/18 G08B13/196

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G07G H04N G08B G07F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP,A,0 221 631 (VIDEOSCAN LTD) 13 May 1987 see column 4, line 19 - column 5, line 46; claims; figures 1,2	1-16
A	PATENT ABSTRACTS OF JAPAN vol. 011 no. 370 (P-642) ,3 December 1987 & JP,A,62 140272 (SONY CORP) 23 June 1987, see abstract	1,3-8,16
A	US,A,4 237 483 (CLEVER ERIC C) 2 December 1980 see column 3, line 4 - column 5, line 37 see column 8, line 67 - column 10, line 57 see column 11, line 32 - column 12, line 12; claims 1-7; figures 1-6	1,2,4-8, 11-16

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
7 August 1995	18.08.95

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Guivol, O

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 95/05167

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PATENT ABSTRACTS OF JAPAN vol. 006 no. 032 (E-096) ,26 February 1982 & JP,A,56 149881 (MITSUBISHI ELECTRIC CORP) 19 November 1981, see abstract ---	1,2,4-8, 16
A	US,A,4 994 916 (PSHTISSKY YACOV ET AL) 19 February 1991 see abstract; claims 1,3-5,10-14; figures ---	1,3-8, 12-16
A	US,A,4 991 008 (NAMA DONALD) 5 February 1991 see abstract; figure 3 see column 8, line 23 - column 12, line 2 ---	1-6, 13-16
A	WO,A,92 11614 (KATZ BARRY) 9 July 1992 cited in the application -----	

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter   nal Application No
PCT/US 95/05167

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-0221631	13-05-87	GB-A- 2182224	07-05-87
		JP-A- 62101190	11-05-87
US-A-4237483	02-12-80	NONE	
US-A-4994916	19-02-91	NONE	
US-A-4991008	05-02-91	NONE	
WO-A-9211614	09-07-92	US-A- 5216502	01-06-93
		AU-B- 660005	08-06-95
		AU-A- 9166391	22-07-92
		DE-T- 563303	17-03-94
		EP-A- 0563303	06-10-93
		EP-A- 0663655	19-07-95
		JP-T- 6503437	14-04-94
		OA-A- 9781	15-04-94

56

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/067884 A1

(51) International Patent Classification[7]: **H04N 7/12,** 5/225, 5/228

(21) International Application Number: PCT/IL03/00097

(22) International Filing Date: 6 February 2003 (06.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/354,209 6 February 2002 (06.02.2002) US
PCT/IL02/01042 26 December 2002 (26.12.2002) IL

(71) Applicant *(for all designated States except US)*: **NICE SYSTEMS LTD.** [IL/IL]; Hapnina Street 8, 43107 Ra'anana (IL).

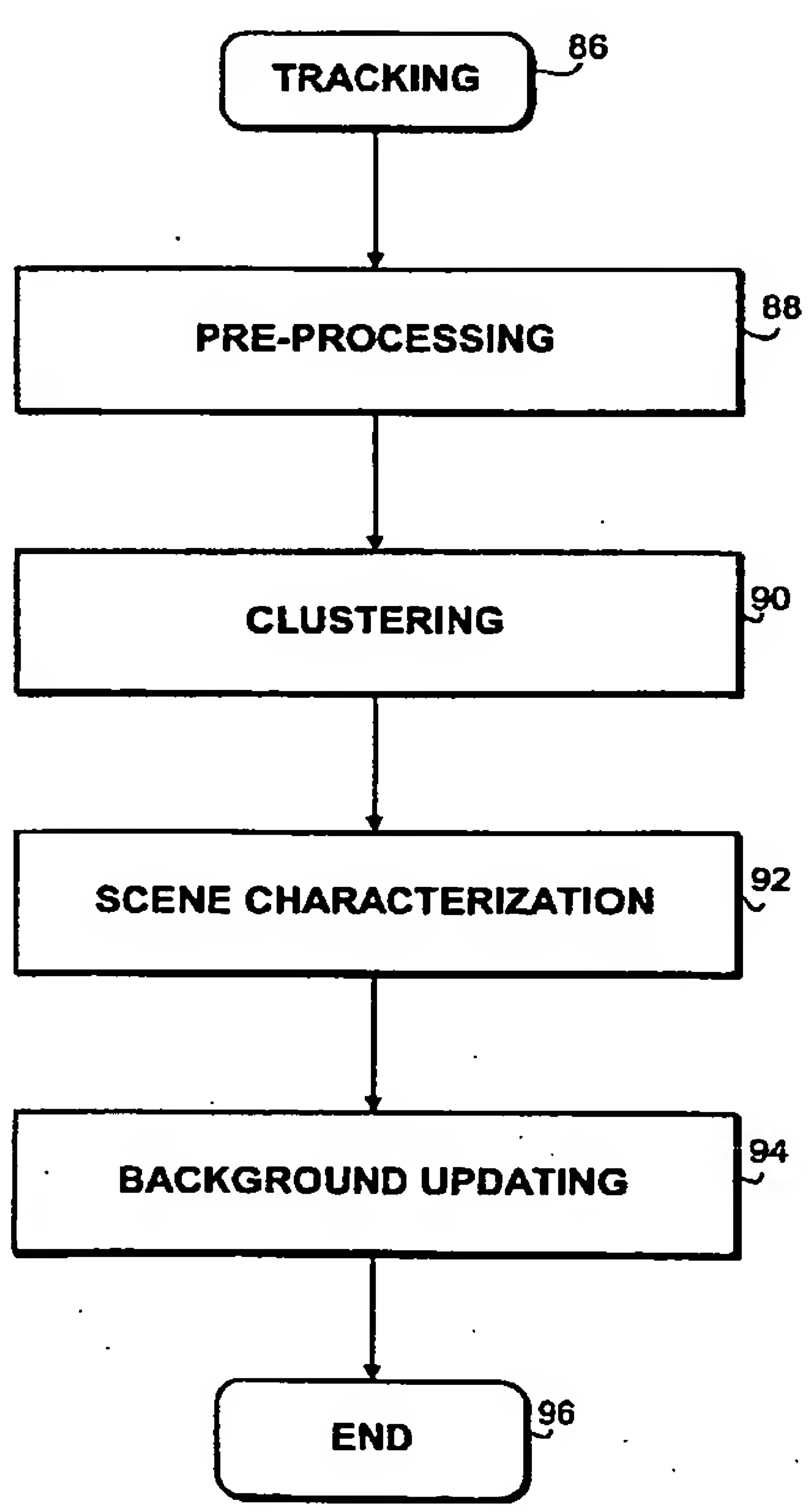
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YEREDOR, Arie** [IL/IL]; Levanon Street 95, 69345 Tel Aviv (IL). **DVIR, Igal** [IL/IL]; Akiva Street 86, 43263 Ra'anana (IL). **KOREN-BLUMSTEIN, Guy** [IL/IL]; Emek Ayalon Street 14, 44627 Kfar Saba (IL). **LACHOVER, Boaz** [IL/IL]; Miriam Hahasmonait 10, 62665 Tel Aviv (IL).

(74) Agents: AGMON, Jonathan et al.; Soroker-Agmon, Advocates and Patent Attorneys, Levinstein Tower, 12th floor, Petach Tikva Road 23, 66184 Tel Aviv (IL).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR VIDEO FRAME SEQUENCE-BASED OBJECT TRACKING



(57) Abstract: An apparatus and method for the analysis of a sequence of captured images covering a scene for detecting and tracking of moving and static objects (86) and for matching (88) the patterns of object behavior in the captured images to object behavior in predetermined scenarios.

WO 03/067884 A1

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND APPARATUS FOR VIDEO FRAME SEQUENCE-BASED OBJECT TRACKING

BACKGROUND OF THE INVENTION

RELATED APPLICATIONS

The present invention relates and claims priority from US provisional patent application serial number 60/354,209 titled ALARM SYSTEM BASED ON VIDEO ANALYSIS, filed 6 February 2002. The present invention also claims priority from and is related to PCT application serial number PCT/IL02/01042 titled SYSTEM AND METHOD FOR VIDEO CONTENT-ANALYSIS-BASED DETECTION, SURVELLANCE, AND ALARM MANAGEMENT, filed 24 December 2002.

FIELD OF THE INVENTION

The present invention relates to video surveillance systems in general, and more particularly to video frame sequence-based objects tracking in video surveillance environments.

DISCUSSION OF THE RELATED ART

Existing video surveillance systems are based on diverse automatic object tracking methods. Object tracking methods are designed to process a captured sequence of temporally consecutive images in order to detect and track objects that do not belong to the "natural" scene being monitored. Current object tracking methods are typically performed by the separation of the objects from the background (by delineating or segmenting the objects), and via the determination of the motion vectors of the objects across the sequence of frames in accordance with the spatial transformations of the tracked objects. The drawbacks of the current methods concern the inability to track static objects for a lengthy period of time. Thus, following a short interval, during which a previously dynamic object ceased moving, the tracking of the same object is effectively rendered. An additional drawback of the current methods concerns the inability of the methods to handle "occlusion" situations, such as where the tracked objects are occluded (partially or entirely) by other objects

temporarily passing through or permanently located between the image acquiring devices and the tracked object.

There is a need for an advanced and enhanced surveillance, object tracking and identification system. Such a system would preferably automate the procedure concerning the identification of an unattended object. Such a system would further utilize an advanced object tracking method that would provide the option of tracking a non-moving object for an operationally effective period and would continue tracking objects in an efficient manner even where the tracked object is occluded.

SUMMARY OF THE PRESENT INVENTION

One aspect of the present invention regards an apparatus for the analysis of a sequence of captured images covering a scene for detecting and tracking of moving and static objects and for matching the patterns of object behavior in the captured images to object behavior in predetermined scenarios. The apparatus comprises at least one image sequence source for transmitting a sequence of images to an object tracking program, and an object tracking program. The object tracking program comprises a pre-processing application layer for constructing a difference image between a currently captured video frame and a previously constructed reference image, an objects clustering application layer for generating at least one new or updated object from the difference image and an at least one existing object, and a background updating application layer for updating at least one reference image prior to processing of a new frame.

A second aspect of the present invention regards a method for the analysis of a sequence of captured images showing a scene for detecting and tracking of at least one moving or static object and for matching the patterns of the at least one object behavior in the captured images to object behavior in predetermined scenarios. The method comprises capturing at least one image of the scene, pre-processing the captured at least one image and generating a short term difference image and a long term difference image, clustering the at least one

moving or static object in the short term difference and long term difference images, and generating at least one new object and at least one existing object.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the drawings in which:

Fig. 1 is a schematic block diagram of the system architecture, in accordance with a preferred embodiment of the present invention;

Fig. 2 is a high-level block diagram showing the application layers of the object tracking apparatus, in accordance with the preferred embodiment of the present invention;

Fig. 3 is a block diagram illustrating the components of the configuration layer, in accordance with the preferred embodiment of the present invention;

Fig. 4A is a block diagram illustrating the components of the pre-processing layer, in accordance with the preferred embodiment of the present invention;

Fig. 4B is a block diagram illustrating the components of the clustering layer, in accordance with the preferred embodiment of the present invention;

Fig. 5A is a block diagram illustrating the components of the scene characterization layer, in accordance with the preferred embodiment of the present invention;

Fig. 5B is a block diagram illustrating the components of the background update layer, in accordance with the preferred embodiment of the present invention;

Fig. 6 is a block diagram showing the data structures associated with the object tracking apparatus, in accordance with a preferred embodiment of the present invention;

Fig. 7 illustrates the operation of the object tracking method, in accordance with the preferred embodiment of the present invention;

Fig. 8 describers the operation of the reference image learning routine, in accordance with a preferred embodiment of the present invention;

Fig. 9 shows the input and output data structures associated with the pre-processing layer, in accordance with a preferred embodiment of the present invention;

Figs. 10A, 10B and 10C describe the operational steps associated with the clustering layer, in accordance with the preferred embodiment of the present invention;

Fig. 11 illustrates the scene characterization, in accordance with the preferred embodiment of the present invention;

Fig. 12 illustrates the background updating, in accordance with the preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

An object tracking apparatus and method for the detection and tracking of dynamic and static objects is disclosed. The apparatus and method may be utilized in a monitoring and surveillance system. The surveillance system is operative in the detection of potential alarm situation via a recorded surveillance content analysis and in the management of the detected unattended object situation via an alarm distribution mechanism. The object tracking apparatus supports the object tracking method that incorporates a unique method for detecting, tracking and counting objects across a sequence of captured surveillance content images. Through the operation of the object tracking method the captured content is analyzed and the results of the analysis provide the option of activating in real time a set of alarm messages to a set of diverse devices via a triggering mechanism. In order to provide the context in which the object tracking apparatus method is useful, several exemplary associated applications will be briefly described. The method of the present invention may be implemented in various contexts such as the detection of unattended objects (luggage, vehicles or persons), identification of vehicles parking or driving in restricted zones, access control of persons into restricted

zones, prevention of loss of objects (luggage or persons) and counting of persons, as well as in police and fire alarm situations. In likewise manner the object tracking apparatus and method described here in may be useful in myriad of other situations and as a video objects analysis tool.

In the preferred embodiments of the present invention, the monitored content is a stream of video images recorded by video cameras, captured, and sampled by a video capture device and transferred to a video processing unit. Each part of this system may be located in a single device or in separate devices located in various locations and inter-connected by hardwire or via wireless connection over local or wide or other networks. The video processing unit performs a content analysis of the video frames where the content analysis is based on the object tracking method. The results of the analysis could indicate an alarm situation. In other preferred embodiments of the invention, diverse other content formats are also analyzed, such as thermal based sensor cameras, audio, wireless linked cameras, data produced from motion detectors, and the like.

An exemplary application that could utilize the apparatus and method of the present invention concerns the detection of unattended objects, such as luggage in a dynamic object-rich environment, such as an airport or city center. Other exemplary applications concern the detection of a vehicle parked in a forbidden zone, or the extended-period presence of a non-moving vehicle in a restricted-period parking zone. Forbidden or restricted parking zones are typically associated with sensitive traffic-intensive locations, such as a city center. Still applications that could use the apparatus and method include the tracking of objects such as persons involved in various scenario models, such as a person leaving the vehicle away from the terminal, which may equal suspicious (unpredicted) behavioral pattern. In other possible applications of the apparatus and method of the present invention can be implemented to assist in locating lost luggage and to restrict access of persons or vehicles to certain zones. Yet other applications could regard the detection of diverse other objects in diverse other environments. The following description is not meant to be

limiting and the scope of the invention is defined only by the attached claims. Several such applications are described in detail in related PCT patent application serial number PCT/IL02/01042 titled SYSTEM AND METHOD FOR VIDEO CONTENT-ANALYSIS-BASED DETECTION, SURVEILLANCE, AND ALARM MANAGEMENT, filed 24 December 2002, the content of which is incorporated herein by reference.

The method and apparatus of the present invention is operative in the analysis of a sequence of video images received from a video camera covering a predefined area, referred herein below to as the video scene. In one example it may be assumed that the object monitored is a combined object comprising an individual and a suitcase where the individual carries the suitcase. The combined object may be separated into a first separate object and a second separate object. It is assumed that the individual (second object) leaves the suitcase (first object) on the floor, a bench, or the like. The first object remains in the video scene without movement for a pre-defined period of time. It is assumed that the suitcase (first object) was left unattended. The second object exits the video scene. It is assumed that the individual (second object) left the video scene without the suitcase (first object) and is now about leave the wider area around the video scene. Following the identification of the previous sub-events, referred to collectively as the video scene characteristics, the event will be identified by the system as a situation in which an unattended suitcase was left in the security-sensitive area. Thus, the unattended suitcase will be considered as a suspicious object. Consequently, the system of the present invention generates, displays, and or distributes an alarm indication. Likewise, in an alternative embodiment a first object, such as a suitcase or person monitored is already present and monitored within the video scene. Such object can be lost luggage located within the airport. Such object can be a person monitored. The object may merge into a second object. The second object can be a person picking up the luggage, another person to whom the first person joins or a vehicle to which the first person enters. The first object (now merged with the second object) may move from its original position and exist

the scene or move in a prohibited direction so predetermined. The application will provide an indication to a human operator. The indication may be oral, visual or written. The indication may be provided visually to a screen or delivered via communication networks to officers located at the scene or to off-premises or via dry contact to an external device such as a siren, a bell, a flashing or revolving light and the like. An additional exemplary application that could utilize the apparatus and method of the present invention regards a detection of vehicles parked in restricted area or moving in restricted lanes. Airports, government buildings, hotels and other institutions typically forbid vehicles from parking in specific areas or driving in restricted lanes. In some areas parking is forbidden all the time while in other areas parking is allowed for a short period, such as several minutes. The second exemplary application is designed to detect vehicles parking in restricted areas for more than a pre-defined number of time units and generates an alarm when identifying an illegal parking event of a specific vehicle. In another preferred embodiment the system and method of the present invention can detect whether persons disembark or embark a vehicle in predefined restricted zones. Other exemplary applications can include the monitoring of persons and objects in city centers, warehouses, restricted areas, borders or checkpoints and the like.

It would be easily perceived that for the successful operation of the above-described applications an object tracking apparatus and an object tracking method are required. The object tracking method should be capable of detecting moving objects, tracking moving objects and tracking static objects, such as objects that are identified as moving and subsequently identified as non-moving during a lengthy period of time. In order to match the patterns of object behavior in the captured image sequences to the patterns of object behavior in above-described scenarios, the object tracking method should recognize linked or physically connected objects, to be able to recognize the separation of the linked objects, to track the separated objects while retaining the historical connectivity states of the objects. The object tracking apparatus and method should further be able to handle occlusions where the tracked

objects are occluded by one or more separate objects temporarily, semi-permanently or permanently.

Referring to Fig. 1 the image sequence sources 12 are one or more video cameras operating in a security-wise sensitive environment and cover a specific pre-defined visual area that is required to be monitored. The area monitored can be any area preferably in a transportation area including an airport, a city center, a building, and restricted or non-restricted areas within buildings or outdoors. The image sequence sources 12 could include analog devices and/or digital devices. The images provided by the image sequence sources could include normal light, infrared, temperature, or any other form of radiation. The image sequence sources 12 continuously acquire and transmit sequences of video images and provide the images simultaneously to an image sequence display device 20 and to a computing and storage device 15. The display device 20 could be a video terminal, which is operated by a human operator or any other display device including a display device located on a mobile or hand held device. Alarm triggers are generated by the object tracking program 14 installed in the computing and storage device 15 in order to indicate an alarm situation to the operator of the display device 20. The alarm may be generated in the form of an audio or any other indication. The image sequence sources 12 transmit sequences of video images to an object tracking program 14 via suitably wired connections. The images could be provided through an analog interface, a digital interface or through a Local Area Network (LAN) interface or Wide Area Network (WAN), IP, Wireless, Satellite connectivity. The computing and storage device 15 could be an external computing platform, such as a personal computer (PC), a UNIX workstation or a mainframe computer having appropriate processing and storage units or a dedicated hardware such as a DSP based platform. It is contemplated that future hand held devices will be powerful enough to also implement device 15 there within. The device 15 could be also an array of integrated circuits with built-in digital signal processing (DSP) and storage capabilities coupled directly to the image sequence sources 12. The device 15

includes a set of object tracking routines constituting the object tracking program 14 and a set of object tracking control data structures 16. The object tracking program 14 in association with the object tracking control data structures 16 receives the image sequence from the image sequence sources 12, and processes the image sequence in order to detect and to track objects therein. Consequent to the detection of pre-defined spatio-temporal patterns of behavior associated with the tracked objects across the image sequences appropriate alarm triggers are generated and transmitted to the display device 20.

Still referring to Fig. 1 the object tracking program 14 and the associated control data structures 16 could be installed in distinct platforms and/or devices distributed randomly across a Local Area Network (LAN) that could communicate over the LAN infrastructure or across Wide Area Networks (WAN). One example is a Radio Frequency Camera that transmits composite video remotely to a receiving station, the receiving station can be connected to other components of the system via a network or directly. The program 14 and the associated control data structures 16 could be installed in distinct platforms and/or devices distributed randomly across very wide area networks such as the Internet. Various forms of communication between the constituent parts of the system can be used. Such can be a data communication network, which can be connected via landlines or wireless or like communication devices and that can be implemented via TCP/IP protocols and like protocols. Other protocols and methods of communications, such as cellular, satellite, low band, and high band communications networks and devices will readily be useful in the implementation of the present invention. The program 14 and the associated control data structures 16 could be further co-located on the same computing platform or distributed across several platforms for load balancing, redundancy considerations, back-up in the case of equipment failure, and the like. Although on the drawing under discussion only a single image sequence source and a single computing and storage device is shown it will be readily perceived that in a realistic environment a plurality of image sequence sources could be

connected to a plurality of computing and storage devices. Moreover, two image sequence sources each capturing a slightly different scene may provide a stereo image sequence source. Likewise, a multiplexed image sequence source from a plurality of image capturing devices may be used. The object tracking apparatus comprises an object tracking program and associated object tracking control data structures.

Referring now to Fig. 2 which is a high-level block diagram showing the application layers of the object tracking apparatus of the present invention. The object tracking program 14 includes several application layers. Each application layer is a group of logically and functionally linked computer program components responsible for different aspects of the application within the apparatus of the present invention. The object tracking program 14 includes a configuration layer 38, a pre-processing layer 42, and an objects clustering layer 44, a scene characterization layer 46, and a background updating layer 48. Each layer is a computer program executing within the computerized environment shown in detail in association with the description of Fig. 1. The configuration layer 38 is a responsible for the initialization of the apparatus of the present invention in accordance with specific user-defined parameters. The pre-processing layer 42 is operative in constructing difference images between a currently captured video frame and previously constructed reference images. The objective of the objects clustering layer 44 is to generate new and or updated objects from the difference images and the existing objects. The scene characterization layer 46 uses the objects generated by the objects clustering layer 44 to describe the monitored scene. The layer 46 also includes a triggering mechanism that compares the behavior pattern and other characteristics of the objects to pre-defined behavior patterns and characteristics in order to create alarm triggers. The background updating layer 48 updates the reference images for the processing of the next frame. A more detailed description of the structure and functionality of the application layers will be provided herein under in association with the following drawings.

Referring to Fig. 3 shows a block diagram illustrating the components of the configuration layer. The configuration layer 38 comprises a reference image constructor component 50, a timing parameters definer component 52, and a visual parameters definer component 54. The reference image constructor component 50 is responsible for the acquisition of the background model. The reference image is generated in accordance with a pre-defined option. The component 50 includes a current frame capture module 56, a reference image loading module 60, and a reference image learning module 62. In accordance with the pre-selected option the reference image may be created alternatively from; a) a currently captured frame, b) an existing reference image, c) a reference image learning module. The current frame capture module 56 provides a currently captured frame to be used as the reference image. The currently captured frame can be a frame from any camera covering the scene. The reference image loading module 60 provides the option for loading an existing reference image located on file locally or remotely. The user may select the appropriate image from the file and designate it as the reference image. The reference image learning module 62 provides the option that the reference image is generated adaptively learned from a consecutive sequence of captured images. The timing parameters definer component 52 provides time settings information, such as the number of time units to be elapsed before the generation of a trigger on a static object, and the like. The visual parameters definer component 54 provides the option to the user to define the geometry of the monitored scene. The component 54 includes, a camera tilt setting module 64, a camera zoom setting module 65, a region location definition module 66, a region type definition module 67, and an alarm type definition module 68. The module 64 derives the camera tilt in accordance with the measurements taken by a user of an arbitrary object located at different location in the monitored scene. The module 65 defines the maximum, the minimum and the typical size of the objects to be tracked. The region location definition module 66 provides the definition of the location of one or more regions-of-interest in the scene. The region type definition module

67 enables the user to define a region of interest as "objects track region" or "no objects track region". The alarm type definition module 68 defines a region of interest as "trigger alarm in region" or "no alarm trigger in region", in accordance with the definitions of the user.

Referring now to Fig. 4A showing a block diagram illustrating the components of the pre-processing layer, in accordance with the preferred embodiment of the present invention. The pre-processing layer 42 comprises a current frame handler 212, a short-term reference image handler 214, a long-term reference image handler 216, a pre-processor module, a short-term difference image updater 220, and a long-term difference image updater 222. Each module is a computer program operative to perform one or more tasks in association with the computerized system of Fig. 1. The current frame handler 212 obtains a currently captured frame and passes the frame to the pre-processor module 218. The short-term reference handler 214 loads an existing short-term reference image and passes the frame to the pre-processor module 218. The handler 214 could further provide calculations concerning the moments of the short term reference image. The long-term reference handler 216 loads an existing long-term reference image and passes the frame to the pre-processor module 218. The handler 216 could further provide calculations concerning the moments of the long term reference image.

The pre-processor module 218 uses the current frame and the obtained reference images as input for processing. The process generates a new short-term difference image and a new long-term difference image and subsequently passes the new difference images to the short-term reference image updater (handler) 220 and the long-term difference image updater (handler) 222 respectively. Using the new difference images the updater 220 and the updater 222 update the existing short-term reference image and the existing long-term reference image respectively.

Referring now to Fig. 4B showing a block diagram illustrating the components of the clustering layer, in accordance with the preferred embodiment of the present invention. The clustering layer 44 comprises an

object merger module 231, an objects group builder module 232, an objects group adjuster module 234, a new objects creator module 236, an object searcher module 240, a Kalman filter module 242, and an object status updater 254. Each module is a computer program operative to perform one or more tasks in association with the computerized system of Fig. 1. The object merger module 231 corrects clustering errors by the successive merging of partially overlapping objects having the same motion vector for a pre-defined period. The objects group builder 232 is responsible for creating groups of close objects by using neighborhood relations among the objects. The object group adjuster 234 initiates a group adjustment processes in order to find the optimal spatial parameters of each object in a group. The new objects constructor module 236 constructs new objects from the difference images, controls the operation of a specific object location and size finder function and adjusts new objects. The new objects may be construed from the difference images whether existing objects are compared with or where there are no existing objects. For example, when the system begins operation a new object may be identified even if there are no previously acquired and existing objects. The object searcher 240 scans a discarded objects archive in order to attempt to locate recently discarded objects with parameters (such as spatial parameters) similar to a newly created object.

In order to improve accuracy of the tracking and in order to reduce the computing load a Kalman filter module 242 is utilized to track the motion of the objects. The object status updater 254 is responsible for modifying the status of the object from "static" to "dynamic" or from "dynamic" to "static". A detailed description of the clustering layer 44 will be set forth herein under in association with the following drawings.

Referring now to Fig. 5A showing a block diagram illustrating the components of the scene characterization layer, in accordance with the preferred embodiment of the present invention. The scene characterization layer 46 comprises an object movement measurement module 242, an object merger module 244, and a triggering mechanism 246. The object movement

measurement module 242 analyzes the changes in the spatial parameters of an object and determines whether the object is moving or stationary. The object merger module 244 is responsible for correcting errors to objects as a result of the clustering stage. The functionality of the triggering mechanism 246 is to check each object against the spatio-temporal behavior patterns and properties defined as "suspicious" or as alarm triggering. When a suitable match is found the mechanism 246 generates an alarm trigger. The operation of the scene characterization layer 46 will be described herein under in association with the following drawings.

Referring now to Fig. 5B showing a block diagram illustrating the components of the background update layer, in accordance with the preferred embodiment of the present invention. The background updating layer 48 comprises a background draft updater 248, a short-term reference image updater 250, and a long-term reference image updater 252. The functionality of the updater 248 is to update continuously the background or reference "draft" frame from the current frame. The short-term reference image updater 250 and the long-term reference image updater 252 maintain the short-term reference image and the long-term reference image, respectively. A detailed description of the operation of the background-updating layer 48 will be provided herein under in association with following drawings.

Referring now to Fig. 6 showing a block diagram of the data structures associated with the object tracking apparatus, in accordance with a preferred embodiment of the present invention. The object tracking control structures 16 of Fig. 1 comprise a long-term reference image 70, a short-term reference image 72, an objects table 74, a sophisticated absolute distance (SAD) short-term map 76, a sophisticated absolute distance (SAD) long-term map 78, a discarded objects archive 82, and a background draft 84. The long-term reference image 70 includes the background image of the monitored scene without the dynamic and without the static objects tracked by the apparatus and method of the present invention. The short-term reference image 72 includes the scene background image and the static objects tracked by the object

tracking method. The objects table includes a list of dynamic and static objects with associated object data and object meta data. The object data includes object identification, objects status, and various control fields, such as a non-moving counter, non-moving-time counter, and the like. The meta data comprises information concerning the current spatial parameters, the properties and the motion vector data of the objects acquired from the previously performed processing on a succession of previous frames. The short-term and long-term sophisticated difference maps (SADs) 76, 78 represent the difference between a currently captured frame and the short-term and long-term reference images 78, 80. The discarded object archive 82 stores discarded objects for object history. The background draft 84 (also referred to as the reference image, but not the short-term or long-term reference images) is a constantly changing image of the monitored scene. Each pixel within each current frame is taken into consideration when calculating the background draft 84. The draft 84 is used for inserting "static" objects to the short-term reference image 72. The background draft 84 constantly reviews the scene background. If an object enters the monitored scene, such object is inserted into the background draft 84. When the method determines that the object is a "static" object (after the object was perceived as stationary across a pre-defined number of captured frames) the pixels of the object are copied from the background draft 84 to the short-term reference image 72.

Referring now to Fig. 7, the object tracking module operates by detecting objects across a temporally ordered sequence of consecutively captured images where the objects do not belong to the "natural" or “static” monitored scene. The object tracking module operates through the use of a central processing unit (not shown) utilizing data structures (not shown). The data structures are maintained on one or more memory or storage devices installed across a hardware environment supporting the application. Fig. 7 illustrates the various steps in the operation of the object tracking method. The configuration step (not shown) is performed prior to the beginning of the tracking (steps 88 through 94). In the configuration step the object-tracking

module is provided with reference images, with timing parameters and with visual parameters, such as regions-of-interest definitions. The provided information enables the method to decide which regions of the frame to work on and in which regions should an alert situation be produced. The configuration step optionally includes a reference image learning step (not shown) in which the background image is adaptively learned in order to construct a long-term and a short-term reference image from a temporally consecutive sequence of captured images. When no stationary objects were detected in the last frames the long-term reference image is copied and maintained as a short-term reference picture. The long-term reference image contains no objects while the short-term reference image includes static objects, such as objects that have been static for a pre-defined period. In the preferred embodiment of the invention, the length of the pre-defined period is one minute while in other preferred embodiments other time values could be used. The long-term reference image and the short-term reference image are updated for background changes, such as changes in the illumination artifacts associated with the image (lights or shadows or constantly moving objects (such as trees) and the like). The video frame pre-processing phase 88 uses a currently captured frame and the short-term and long-term reference images for generating new short-term and long-term difference images. The difference images represent the difference between the currently captured frame and the reference images. The reference images can be obtained from one of the image sequence sources described in association with Fig. 1 or could be provided directly by a user or by another system associated with the system of the present invention. The difference images are suitably filtered or smoothened. The clustering phase 90 generates new or updated objects from the difference images and from the previously generated or updated objects. The scene characterization phase 92 uses the objects received from the clustering phase 90 in order to describe the scene. The background updating step 94 updates the short-term and long-term reference images for the next frame calculation. Note should be taken that in other preferred embodiments of the invention other

similar or different processes could be used to accomplish the underlying objectives of the method of the present invention.

Note should be taken that proposed apparatus and method is provided the capability of functioning in specific situations where an image acquiring device, such as a video camera, is not static. Examples for such situations include a pole-mounted outdoor camera operating in windy conditions or mobile a camera physically tacking a moving object. For such situations the object tracking method requires a pre-pre-processing phase configured such as to compensate for the potential camera movements between the capture of the reference images and the capture of each current frame. The pre-pre-processing phase involves an estimation of the relative overall frame movement (registration) between the current frame and the reference images. Consequent to the estimation of the registration (in terms of pixel offset) the offset is applied to the reference images in order to extract “in-place” reference images for the object tracking to proceed in a usual manner. As a result, extended reference images have to be used, allowing for margins (the content of which may be constantly updated) up to the maximal expected registration.

The estimation of the registration (offset) between the current frame and the reference images involves a separate estimation of the x and y offset components, and a joint estimation of the x and y offset components. For the separate estimation, selected horizontal and vertical stripes of the current frame and the reference images are averaged with appropriate weighting, and cross-correlated in search of a maximum match in the x and y offsets, respectively. For the joint estimation, diagonal stripes are used (in both diagonal directions), from which the x and y offsets are jointly estimated. The resulting estimates are then averaged to produce the final estimate.

Referring now to Fig. 8 which describers the operation of the reference image learning routine, in accordance with a preferred embodiment of the present invention. The construction of the long-term and short term reference images could be carried out in several alternative ways. A currently captured frame could be stored on a memory device as the long-term reference

image. Alternatively, a previously stored long-term reference image could be loaded from the memory device in order to be used as the current long-term reference image respectively. Alternatively, a specific reference image learning process could be activated (across steps 100 through 114). In step 100 the reference image learning process is performed across a temporally consecutive sequence of captured images where each of the frames is divided into macro blocks (MB) having a pre-defined size, such as 16X16 pixels or 32X32 pixels or any like other division into macro blocks. Next at step 102 each MB is examined for motion vectors. The motion is detected by comparing the MB in a specific position in currently captured frame to the MB in the same position in the previously captured frame. The comparison is performed during the encoding step by using similar information generated therein for video data compression purposes. According to the result of the examination each MB is marked as being in one of the following three states; a) Motion MB 108 where a motion vector is detected in the current MB relative to the parallel MB in the previously captured frame, b) Undefined MB 104 where no motion vector is detected in the MB relative to the parallel MB in the previously captured frame but motion vector was detected across a previously captured set of temporally consecutive frames where the sequence is defined as having a pre-defined number of frames. In the preferred embodiment of the invention the number of frames in the sequence is about 150 frames while in other preferred embodiments of the invention different values could be used, c) Background MB 106 where no motion vector was detected across the previously captured sequence of temporally consecutive frames. In step 110 the values of each of the pixels in an MB that were identified as a Background MB are obtained and in step 112 the values are averaged in time 112. In step 114 an initial short term and long term reference image is generated from the values average in time. In order to avoid undetermined values for pixels in the MBs that were always in motion, such as an MB wherein there was a constant motion (trees moving in wind), in step 114 the short-term reference image is created such that it contains the averages of the values of pixels in time. Subsequently, the pixels

are examined in order to find which pixels had insufficient background time (MBs that were always in motion). Pixels without sufficient background time are given the value from the short-term reference image.

Referring now to Fig. 9 showing the input and output data structures associated with the pre-processing layer, in accordance with a preferred embodiment of the present invention. The pre-processing step 88 of Fig. 6 employs the current frame 264 and the short-term reference image 262 to generate a short-term difference image 270. The step 88 further uses the current frame 264 and the long-term reference image 266 to generate a long-term difference image 272. The long-term 272 and short-term 270 difference images represent respectively the sophisticated absolute difference (SAD) between the current frame 264 and the long-term 266 and the short-term 262 reference images. The size of the difference images (referred to herein after as SAD maps) 270, 272 is equal to the size of the current frame 264. Each pixel in the SAD maps 270, 272 are provided with an arbitrary value in the range of 1 through 101. Other values may be used instead. High values indicate a substantial difference between the value of the pixel in the reference images 262, 266 and the value of the pixel in the currently captured frame 264. Thus, the score indicates the probability for the pixel belonging either to the scene background or to an object. The generation of the SAD maps 270, 272 is achieved by performing one of two alternative methods.

Still referring to Fig. 9, in the first pre-processing method for each specific pixel in the currently captured frame 264 the absolute difference between the specific pixel and the matching pixel in the reference images 262, 266 is calculated where the calculation takes into account the average pixel value:

(1): $D(x, y) = a0 \times Ymin(x, y) + a1 \times Ymax(x, y) + a3$

In the above equation the values of x, y concern the pixel coordinates. The values of Ymin and of Ymax represent the lower and the higher luminance levels at (x, y) between the current frame 264 and the

reference images 262, 266. The values of a0, a1, and a3 are thresholds designed to minimize D(x, y) for similar pixels and maximize it for non-similar pixels. Consequent to the performance of the above equation for each of the pixels and to the generation of the SAD maps 270, 272 the SAD maps 270, 272 are filtered for smoothing with two Gaussian filters one in the X coordinate and the second in the Y coordinate.

In the second alternative pre-processing method, around each pixel P(x, y) the following values are calculated where the calculation uses a 5X5 pixels neighboring window for filtering. This step could be referred to as calculating the moments of each pixel.

(2):

$$M00(x,y)=\sum_{i=x-2}^{x+2}\sum_{j=y-2}^{y+2}P(i,j)$$

$$M10(x,y)=\frac{32*\sum_{i=x-2}^{x+2}\sum_{j=y-2}^{j=y+2}(x-i)P(i,j)}{M00(x,y)}$$

$$M01(x,y)=\frac{32*\sum_{i=x-2}^{x+2}\sum_{j=y-2}^{y+2}(y-j)P(i,j)}{M00(x,y)}$$

The results of the equations represent the following values: a) M00 is the sum of all the pixels around the given pixel, b) M10 is the sum of all the pixels around the given pixel each multiplied by a filter that detects horizontal edges, and c) M01 is the sum of all the pixels around a given pixel multiplied by a filter that detects vertical edges. Next, the absolute difference between these three values in the current frame 264 and the reference images 270, 272 is performed. In addition the minimum of M00Curr and M00Ref are calculated.

(3):

$$D00(x,y)=|M00curr(x,y)-M00ref(x,y)|$$
$$D10(x,y)=|M10curr(x,y)-M10ref(x,y)|$$
$$D01(x,y)=|M01curr(x,y)-M01ref(x,y)|$$
$$Min(x,y)=\min(M00curr(x,y),M00ref(x,y))$$

Next the following equations are used to construct the desired SAD maps 270, 272:

(4):

$$Tmp1(x,y) = A0*(D00(x,y)+W0) - Min(x,y)$$
$$Tmp2(x,y) = A1*D10(x,y)+W1$$
$$Tmp3(x,y) = A1*D01(x,y)+W1$$
$$A0 = 15,$$
$$A1 = 25$$
$$W0 = -40$$
$$W1 = -44$$

(5):

$$Tmp1(x,y) = \min(32, Tmp1(x,y))$$
$$Tmp1(x,y) = \max(-32, Tmp1(x,y))$$
$$Tmp2(x,y) = \min(32, Tmp2(x,y))$$
$$Tmp2(x,y) = \max(-32, Tmp2(x,y))$$
$$Tmp3(x,y) = \min(32, Tmp3(x,y))$$
$$Tmp3(x,y) = \max(-32, Tmp3(x,y))$$

(6):

$$TmpSADMap(x,y) = \frac{3*(Tmp1(x,y)+Tmp2(x,y)+Tmp3(x,y)+32)}{64}$$

Through a convolution calculation the grade for each pixel is calculated while taking into consideration the values for the pixels neighbors:

(7):

$$SADMap(x,y) = 1 + \sum_{i=x-2}^{x+2} \sum_{j=y-2}^{y+2} TmpSADMap(i,j)$$

$$SADMap(x,y) = \min(SADMap(x,y), 101)$$

The method takes into consideration the texture of the current frame 264 and the reference images 262, 266 and compares there between. The second pre-processing method is favorable since it is less sensitive to light changes

At the price of increased computational cost, in order to achieve a more accurate model optionally higher moments could be calculated.

Calculating higher moments involves the performance of the following set of equations:

(8):

$$M20(x,y) = \frac{\sum_{i=x-2}^{x+2} \sum_{j=y-2}^{y+2} (x-i)^2 * P(i,j)}{M00}$$

$$M02(x,y) = \frac{\sum_{i=x-2}^{x+2} \sum_{j=y-2}^{y+2} (y-j)^2 * P(i,j)}{M00}$$

$$M11(x,y) = \frac{\sum_{i=x-2}^{x+2} \sum_{j=y-2}^{y+2} (y-j) * (x-i) * P(i,j)}{M00}$$

It will be easily perceived that the method could be broadened for even higher moments. Several equations of the second pre-processing method represent a simulation of a neural network.

The pre-processing step produces several outputs that are used by the clustering step. Such outputs are the short-term and long-term SAD maps. Each pixel in the SAD maps is assigned a value in the range of 1 through 100. High values indicate great difference between the value of the pixel in the reference images and the value of the pixel in the current frame. The purpose of the clustering is to cluster the high difference pixels into objects. Referring now to Fig. 10A the clustering step 120 includes a two-stage Kalman filtering, two major processing sections, and an object status updating. In order to improve accuracy of the tracking and in order to reduce the computing load a Kalman filter is used to track the motion of the objects. The Kalman filter is performed in two steps. The prediction step 120 is performed before the adjustment of the objects and the update step 125 is performed after the creation of a new object. The Kalman state of the object is updated in accordance with the adjusted parameters of the object. At step 204 the status of the object is updated. The changing of the object status from "dynamic" status to "static" status is performed as follows: If the value of the non-moving counter associated with the object exceeds a specific threshold then the status of the object is set to

"static". The dead-area (described in the clustering step) is calculated and saved. The pixels that are bounded within the object are copied from the background draft to the short-term reference image. Subsequently, the status of the object is set the "static". Static objects are not adjusted until their status is changed back to "dynamic".

Still referring to Fig. 10A in the processing step 122, in order to perform tracking of the objects that were detected in the previous video frames, the parameters of the existing objects are adjusted. In the processing step 124 new objects are created from all the high value pixels that do no belong to the already created objects. The adjustment of the object parameters is done for every group of objects. The objects are divided into groups in accordance to their location. Objects of a group are close to each other and might occlude each other. Objects from different groups are distant from each other. The adjustment of groups of objects provides for the appropriate handling of occlusion situations.

Referring now to Fig. 10B at step 126 the objects groups are built. An object-specific bounding ellipse represents each object. The functionality, structure and operation of the ellipse will be described herein after in association with the following drawings. Every two objects are identified as neighbors if the minimum distance between their bounding ellipses is up to about 4 pixels. Using the neighborhood relations between every two objects, the object groups are built. Note should be taken that static objects are not adjusted. At step 128 the parameters of the existing dynamic objects are adjusted in order to perform tracking of the objects detected in the previously captured video frames. The objects are divided into groups according to their locations. Objects of a group are close to each other and may occlude each other. Objects belonging to different groups are distant from each other. The adjustment of the object parameters is performed for every group of objects separately. The adjustment to groups of objects enables appropriate handling of occlusion situations. At step 126 groups of objects are built. Each object is represented by a bounding marker, which a distinct artificially generated

graphical structure, suoh as an ellipse. A pair of objects is identified as two neighboring members if the minimum distance between their marker ellipses is up to a pre-defined number of pixels. In the preferred embodiment of the invention the pre-defined number of pixels is 4 while in other embodiments different values could be used. Using the neighborhood relations between all the pairs of objects the object groups are built. At step 128 the object groups are adjusted. The object group adjustment process determines the optimal spatial parameters of each object in the objects group. Each set of spatial parameter values of all the objects in a given objects group is scored. The purpose of the adjustment process is to find the spatial parameters of each object in a group, such that the total score of the group is maximized. The initial parameters are the values generated for the previously captured frame. The initial base score is derived from a predictive Kalman filter. In each adjustment iteration, a pre-defined number of geometric operations are performed on the objects. The operations effect changes in the parameters of every object in the group. Various geometric operations could be used, such as translation, scaling (zooming), rotation, and the like. In the preferred embodiment of the invention, the number of geometric operations applied to the object is 10 while in other preferred embodiments different values could be applied. In the preferred embodiment of the invention, the following geometrical operations with the respective values are used: a) Translation right on axis 1, b) Translation left on axis 1, c) Translation right on axis 2, d) Translation left of axis 2, e) Down-scaling by shrinking axis 1, f) Up-scaling by blowing axis 1, g) Down-scaling by shrinking axis 2, h) Up-scaling by blowing axis 2, i) Rotation to the left through 5 degrees, and j) Rotation to the right through 5 degrees. The score of every change is measured and saved in a table. The structure and the constituent elements of the table are described via a representation of an exemplary table as follows:

Adj 1	Adj 2	Adj 3	Adj 4	Adj 5	Adj 6	Adj 7	Adj 8	Adj 9	Adj 10
100	102	101	105	104	108	110	108	100	120

150	80	105	104	110	112	114	121	119	120
123	121	112	114	119	117	109	108	105	101

In the example above there are 3 objects in the group where each row represents an object. The 10 adjustments performed on each object are represented by the results shown in each row. Performing adjustment 1 to the 2^{nd} object yields the maximum score for the group. Thus, adjustment 1 will be applied to the parameters of the 2^{nd} object. The score is weighted by the non-movement-time of the object. As a result the algorithm tends not to perform changes on objects that were not in movement for a significant period. The iterative process is performed in order to improve the score of the group as much as possible. The iterative process stops if at last one of the following conditions is satisfied: a) the highest score found in the iteration is no greater than the score at the beginning of the iteration, and b) at least twenty iterations have been completed.

In order to reduce the computational load, every ellipse parameter is changed according to the movement thereof as derived by a Kalman filter used to track after the object. If the score of the group is higher than the base score the change is applied and the new score will become the base score.

In order to handle occlusions a "united object" is built, which is a union of all the objects in the group. Thus, each pixel that is associated with more than one object in the group, will contribute its score only once and not for every member object that wraps it. The contribution of each pixel in the SAD map to the total score of the group is set in accordance with the value of the pixel.

(9):

$$Contribution = \begin{cases} +2 & HighTH < val \\ +1 & LowTH < val < HighTH \\ -1 & val < LowTH \end{cases}$$

Subsequent to the completion of the about 10 iterations, specific object parameters associated with each group object are tested against specific

thresholds in order to check whether the object should be discarded. The object parameters to be tested are the minimum object area, the minimum middle score, the maximum dead area, and the overlap ratio.

a) Maximum object area concerns a threshold value limiting the minimum permissible spatial extent for an object. If the maximum object area is smaller than the value of a pre-defined threshold then the object is discarded. So for example random appearance of non-real objects or dirty lenses providing random dark pixels are cleaned effectively.

b) Minimum middle score relates to the score of a circle that is bounded in the ellipse representing the object. If the score of the circle is below a pre-defined value of an associated threshold then the object is eliminated. A low-score circle indicates a situation where two objects were in close proximity in the scene and thus represented on object (one ellipse) and then they separated. Thus, the middle of the ellipse will have a lower score than the rest of the area of the object.

c) Maximum dead area concerns an object that includes a large number of low value pixels. If the number of such pixels is higher than an associated threshold value then the object is discarded.

d) Overlap ratio regards occlusion situations. Occlusion is supported up to about 3 levels. If most of the object is occluded by about 3 other objects for a period of about 10 seconds, the object is a candidate to be discarded. If there is more than one object in that group that should be eliminated then the most recently moving object is discarded.

Subsequent to the completion of the parameters testing procedure the non-discarded objects are cleared from the SAD map by setting the value of the set of pixels bounded in the object ellipse to zero. The discarded objects are saved in the discarded objects archive to be utilized as object history. The data of every new object will be compared against the data of the recently discarded objects stored in the archive in order to provide the option of restoring the object from the archive.

Referring now to Fig. 10C consequent to the adjustment of the existing objects, the pixels in the SAD map are provided with values in the range of 0 through 100. A value of zero means that the pixel belongs to an existing object. The drawing shows the steps in the creation of new objects. The construction of a new object is based a pixel having a high value in the SAD map. The procedure starts by searching for a free entry in the objects table 74 of Fig. 5 in order to enable the storage of the parameters of a new object (not shown). The high value pixel is assumed to be the center of the object. In order to derive the boundary of the new object a specific boundary locater function, referred to herein after as the "spider function" is activated at step 130. The spider function includes a set of program instructions associated with a control data structure. The control data structure contains location and size data that define the spatial parameters of a spider-like graphical structure. The spider-like structure is provided with about 16 extensible members (arms) uniformly divided across 360 degrees. The extensible members of the spider-like structure are connected to the perceived center of the new object and dynamically radiate outward. The length of each extensible member is successively increased until the far end of spatially each member is aligned with a pixel having a high value in the SAD map. In order to handle small gaps in the object "bridging" line segments of up to 4 pixels are allowed. Thus, if there are more than 4 continuous low value pixels in the direction of the radiation, the extension of a member will be discontinued. The member-specific final coordinates are saved in X, Y arrays in the control data structure, respectively, in order to indicate the suitable boundary points constituting the boundary line of the new object. Next, in order to improve accuracy the central point of the spider structure is re-calculated from the X, Y arrays, as follows: (10):

$$y_c = \frac{1}{16}\sum_{k=0}^{15} Y[k] \qquad x_c = \frac{1}{16}\sum_{k=0}^{15} X[k]$$

Then, subsequent to the re-location the central point of the structure at the Yc, Xc pixel coordinates the spider structure is re-built. Extending the

about 16 extensible members of the spider structure yields two Y[16] and X[16] arrays. If the spatial extent of the spider structure is sufficient the parameters of the boundary ellipse are calculated. If the spatial extent of the spider overlaps the area of an existing object the new object will not be created unless its size is above a minimum threshold.

Still referring to Fig. 10C at step 132 the spider-like graphical structure is converted to an ellipse-shaped graphical structure. An ellipse is provided with 5 parameters calculated from the X, Y arrays as follows:

(11):

$$M_x = \tfrac{1}{16}\sum_{k=0}^{15} X[k] \qquad M_y = \tfrac{1}{16}\sum_{k=0}^{15} Y[k]$$

$$C_{xx} = \tfrac{1}{16}\sum_{k=0}^{15}(X[k]-M_x)^2 \qquad C_{xy} = \tfrac{1}{16}\sum_{k=0}^{15}(X[k]-M_x)(Y[k]-M_y) \qquad C_{yy} = \tfrac{1}{16}\sum_{k=0}^{15}(Y[k]-M_y)^2$$

The covariance matrix of the ellipse is:

(12):

$$C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{xy} & C_{yy} \end{bmatrix}$$

The ellipse covariance matrix is scaled to wrap the geometric average distance. The covariance matrix is multiplied by where F is calculated in the following manner

(13):

$$d_k = \begin{bmatrix} X[k] & Y[k] \end{bmatrix} \cdot C^{-1} \cdot \begin{bmatrix} X[k] \\ Y[k] \end{bmatrix} \qquad k = 0..15$$

$$F = \left(\prod_{k=0}^{15} d_k \right)^{1/16}$$

At step 134 the new object is adjusted via the utilization of the same adjustment procedure used for adjusting existing objects. The discarded objects archive includes recently discarded objects. If the spatial parameters, such as location and size, of a recently discarded object are similar to the parameters of the new object, the discarded object is retrieved from the archive and the

tracking thereof is re-initiated. If no similar object is found in the archive then the new object will get a new object ID, and the new object's data and meta data will be inserted into the objects table. Subsequently tracking of the new object will be initiated.

Referring now to Fig. 11 the output of the clustering step is the updated spatial parameters of the object stored in the object table. The scene characterization layer 208 uses the existing objects to describe the scene. The layer 208 includes program sections that analyze the changes in the spatial parameters of the object, characterize the spatio-temporal behavior pattern of the object, and update the properties of the object. The temporal parameters and the properties of the object are suitably stored in the objects table. At step 210 object movement is measured. The measurement of the object is performed as follows:

(14):

$$dV_x = \mathrm{sgn}(MeanX - PrevMeanX) \qquad dV_y = \mathrm{sgn}(MeanY - PrevMeanY)$$
$$AccMoveX = 0.5 \cdot AccMoveX + 0.5 \cdot dV_x \qquad AccMoveY = 0.5 \cdot AccMoveY + 0.5 \cdot dV_y$$
$$AccDist = \sqrt{AccMoveX^2 + AccMoveY^2}$$

MeanX/Y is the location of the center of the object in the current frame. PrevMeanX/Y is the location of the center of the object in the previous frame. The value of non-moving counter is updated in accordance with AccDist as follows:

(15):

$$NonMoveCnt = \begin{cases} 0.95 \cdot NonMoveCnt & AccDist > 0.8 \\ NonMoveCnt + 1 & otherwise \end{cases}$$

In the unattended luggage application there is a possibility that a standing or sitting person that does not make significant movements will generate an alarm. In order to handle such false alarms, the algorithm checks whether there is motion inside the object ellipse. If in at least 12 of the last 16 frames there was motion in the object, it is considered as a moving object.

Consequently, the value of the non-moving counter is divided by 2. At step 212 an object merging mechanism is activated. There are cases in which an element in the monitored scene, such as a person or a car, is represented by 2 objects whose ellipses are partially overlapping due to clustering errors. The object merging mechanism is provided for the handling of the situation. Thus, for example, if at least 2 objects are close enough to each other; ("close" as defined in for the clustering process) and are moving with the same velocity for more than 8 frames then the two objects are considered as representing the same element. Thus, the objects will be merged into a single object and a new ellipse will be created to bound the merged objects. The new ellipse data is saved as the spatial parameters of the older object while the younger object is discarded. Each merge is performed between 2 objects at a time. If there are more than 2 overlapping objects that move together additional merges will be performed. Following the characterization of each object's spatio-temporal behavior pattern and other properties, such as texture (including but not limited to color), shape, velocity, trajectory, and the like, against the pre-defined behavior patterns and properties of "suspicious" objects, at step 214 the objects whose behavior pattern and properties are similar to the "suspicious" behavior and properties will generate an alarm trigger. Note should be taken that the suspicious behavior patterns and suspicious properties could vary among diverse applications.

Referring now to Fig. 12 the background update layer updates the reference images for the next frame calculation. The method uses two reference images: a) the long-term reference image, and b) the short-term reference image. The long-term reference image describes the monitored scene as a background image without any objects. The short-term reference image includes both the background image and static objects. Static objects are defined as objects that do not belong to the background, and are non-moving in the monitored scene for a pre-defined period. In the preferred embodiment of the invention the pre-defined period is defined as having a length of about 1 to 2 minutes. In other embodiments different time unit values could be used. The

background updating process uses the outputs of all the previous layers to generate a new short-term reference image. Each pixel that satisfies the following conditions is updated: a) similar enough to the short-term reference image (according to the score given in the pre-processing step), and b) not included in an object. Pixels that do not satisfy the first condition but satisfy the second condition for a long sequence of frames get updated as well. For every fixed number of frames, a comparison is made between the current reference images to the previous reference images, in order to check if the changes made to the reference images were correct. The long-term reference image is updated from the short-term reference image in all pixels that are not contained in any of the tracked objects. An object may change its status from dynamic to static if it is not moving for a given period. It can change its status from static to dynamic if the score thereof in the long-term reference image significantly decreases. The background maintenance could be augmented by user-initiated updates. Thus, the user can add several objects to the background in order to help the system overcome changes in the background due to changes in the location of a background object. For example a "bench" object that was dragged into the scene will be identified by the method as an object. The user can classify the object as a neutral object and therefore can add the object to the background in order to prevent the identification thereof as a dynamic or a static object.

Still referring to Fig. 12 at step 198 the background draft frame is updated. The background draft frame is continuously updated from the current frame in all macro-blocks (16 X 16 pixels or the like) in which there was mo motion for several frames. Each pixel in the background draft is updated by utilizing the following calculation:

(16):

Background Draft (x, y) = Background Draft (x, y) + sgn (Current Frame (x, y) – Background Draft (x, y)

When an object is identified as a static object, it is assumed that the identified object already appears in the background draft. Thus, the pixels of

the object are copied from the background draft to the short-term reference image. The short-term reference image is updated at step 200. The update of each pixel in short-term reference image is performed in accordance with the values of the pixel in the SAD map and in the objects map. In the update calculations the following variables are used:

SAD (x, y) = the SAD map value in the x, y pixel location

OBJECT(x, y) = the number of objects that the pixel in the x, y location belongs to

BACKGROUND_COUNTER (x, y)

NOT_BACKGROUND_COUNTER (x, y)

The previously defined counters are updated by performing the following sequence of instructions:

If (SAD (x, y) < 50) and (OBJECT (x, y) = 0) then the according to the SAD map the pixel belongs to the background and does not belong to any object. Therefore, the value of the BACKGROUND_COUNTER (x, y) is incremented by one. If SAD (x, y) >50 and (OBJECT (x, y) = 0) then the pixel does not belong to the background and does not belong to any object. Therefore, the value of the NOT_BACKGROUND_COUNTER is incremented by one. If OBJECT (x, y) not equal to 0 then there is at least one object that the pixel belongs to. Thus both counters are set to zero. Consequent to the updating of the counters the pixels are updated in accordance with the counters. If BACKGROUND_COUNTER (x, y) greater than or equal to 15 then the pixel at the x, y coordinates is updated and the counter is set to zero. If NOT_BACKGROUND_COUNTER (x, y) greater than or equal to 1000 then the pixel at the x, y coordinates is updated and counter se to zero.

At step 202 the long-term reference image is updated by copying all the pixels that are not bounded by any object's ellipse from the short-term reference image to the long-term reference image.

In the short-term reference image the score of each static object is measured. The score are compared to the score obtained when the object became static. If the current score is significantly lower than the previous score

it is assumed that the static object has started moving. The status of the object is set to "dynamic" and the pixels of the object are copied from the long-term reference image to the short-term reference image. Thus, the object will be adjusted for the next frame during the adjustment process.

The applications that could utilize the system and method of object tracking will now be readily apparent to person skilled in the art. Such can include crowd control, people counting, an offline and online investigation tools based on the events stored in the database, assisting in locating lost luggage (lost prevention) and restricting access of persons or vehicles to certain zones, unattended luggage detection, "suspicious" behavior of persons or other objects and the like. The applications are both for city centers, airports, secure locations, hospitals, warehouses, border and other restricted areas or locations and the like.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined only by the claims, which follow.

CLAIMS

WHAT IS CLAIMED IS:

1. An apparatus for the analysis of a sequence of captured images covering a scene for detecting and tracking of moving and static objects and for matching the patterns of object behavior in the captured images to object behavior in predetermined scenarios, the apparatus comprising the elements of:

   at least one image sequence source for transmitting a sequence of images to an object tracking program; and

   an object tracking program comprising;

   a pre-processing application layer for constructing a difference image between a currently captured video frame and a previously at least one constructed reference image showing the background;

   an objects clustering application layer for generating at least one new or updated object from the difference image; and

   a background updating application layer for updating at least one reference image prior to processing of a new frame.

2. The apparatus of claim 1 wherein the object tracking program further comprises a configuration application layer for initializing the apparatus in accordance with user pre-defined parameters.

3. The apparatus as claimed in claim 2 wherein the configuration application layer comprises a reference image constructor, the reference image constructor comprising a current frame capture module for assigning a captured image as the reference image.

4. The apparatus as claimed in claim 2 wherein the configuration application layer comprises a reference image constructor, the reference image constructor comprising a reference image loading module for loading an existing reference image located on file as the reference image.

5. The apparatus as claimed in claim 2 wherein the configuration application layer comprises a reference image constructor, the

reference image constructor comprising a reference image learning module for generating a reference image from a consecutive sequence of captured images.

6. The apparatus as claimed in claim 2 wherein the configuration application layer comprises a timing parameters definer for providing time setting information.

7. The apparatus as claimed in claim 2 wherein the configuration application layer comprises the element of a visual parameters definer, the visual parameters definer for providing the geometry of the scene.

8. The apparatus as claimed in claim 7 wherein the visual parameters definer comprises a camera tilt setting module for deriving camera tilt in accordance with measurements of an object located at different locations in the scene.

9. The apparatus as claimed in claim 7 wherein the visual parameters definer comprises a camera zoom setting module for defining the maximum, the minimum and the typical size of the objects to be tracked.

10. The apparatus as claimed in claim 7 wherein the visual parameters definer comprises a region location definition module for defining the location of at least one region-of-interest within the scene.

11. The apparatus as claimed in claim 7 wherein the visual parameters definer comprises a region type definition module for defining a region of interest in the scene.

12. The apparatus as claimed in claim 7 wherein the visual parameters definer comprises an alarm type definition module for defining a region of interest as a trigger alarm region.

13. The apparatus as claimed in claim 1 wherein the pre-processing application layer comprises:

a current frame handler for obtaining a captured frame;

a short term reference image handler for loading an existing short-term reference image;

a long term reference image handler loads an existing long-term reference image;

a pre-processor module for generating a new short term and long term reference images;

a short term difference image handler for updating the short term reference image with the new short term reference image; and

a long term reference image handler for updating the long term reference image with the new long term reference image.

14. The apparatus of claim 13 wherein the short and long term reference image handlers further provide the moments of the short and long term reference images.

15. The apparatus as claimed in claim 1 wherein the clustering application layer comprises:

an object merger module for correcting clustering errors by successive merging of at least two partially overlapping objects having the same motion vector for a pre-defined period of time;

an objects group builder module for creating at least one group of at least two close objects;

an object group adjuster module for determining the spatial parameters of each object in the at least one group; and

a new objects constructor module for constructing a new object based on the difference image.

16. The apparatus as claimed in claim 15 wherein the clustering application layer further comprises an object searcher module for locating discarded objects having spatial parameters similar to the parameters of the new object.

17. The apparatus as claimed in claim 15 wherein the clustering application layer further comprises a Kalman filtering module;

18. The apparatus as claimed in claim 15 wherein the clustering application layer further comprises an object status updater module for modifying the status of an object.

19. The apparatus as claimed in claim 1 wherein the objects clustering application layer generates at least one new or updated object from the difference image and an at least one existing object.

20. The apparatus of claim 1 wherein the object tracking program further comprises a scene characterization application layer for describing the scene and for triggering an alarm, based on comparing a behavior pattern of the at least one existing object to the at least one pre-defined behavior pattern or characteristic.

21. The apparatus as claimed in claim 20 wherein the scene characterization application layer comprises an object movement measurement module for analyzing changes in the parameters of the at least one existing object and determining the at least one existing object movement.

22. The apparatus as claimed in claim 20 wherein the scene characterization application layer comprises an object merger module for correcting errors the at least one existing object and an alarm triggering mechanism for determining whether an alarm is to be triggered based on the at least one existing object patterns.

23. The apparatus claimed in claim 1 wherein the background update application layer comprises a background draft updater module for updating the at least one reference image from the currently captured video frame.

24. The apparatus claimed in claim 23 wherein the background update application layer further comprises a short term reference image updater module and a long term reference image updater module for maintaining the updated short term and long term reference images.

25. The apparatus claimed in claim 1 further comprising an object tracking control database, the database comprising;

at least one long term reference image, the at least one long term reference image comprising a background image of the scene without dynamic or static objects tracked by the apparatus;

a short term reference image, the at least one short term reference image comprising a background image of the scene with the dynamic or static objects tracked by the apparatus.

26. The apparatus claimed in claim 25 wherein the object tracking control database further comprising;

an objects table comprising a list of the dynamic or static objects tracked by the apparatus, each object is associated with object data and object meta data; and

a distance short term map and a distance long term map showing the short-term and long-term reference images; and

a background draft comprising a changing image of the scene and making up the reference image.

27. The apparatus claimed in claim 26 wherein the object tracking control database further comprising a discarded objects archive for storing discarded objects.

28. A method for the analysis of a sequence of captured images showing a scene for detecting and tracking of at least one moving or static object and for matching the patterns of the at least one object behavior in the captured images to object behavior in predetermined scenarios, the method comprising the step of:

capturing at least one image of the scene;

pre-processing the captured at least one image and generating a short term difference image and a long term difference image;

clustering the at least one moving or static object in the short term difference and long term difference images and generating at least one new object and at least one existing object.

29. The method as claimed in claim 28 further comprising the steps of characterizing the visual scene and updating the background reference image by updating the short term reference frame and the long term reference frame.

30. The method as claimed in claim 28 further comprising the step of configuring the object tracking program for providing at least one reference image, at least one timing parameter and at least one visual parameter.

31. The method as claimed in claim 28 further comprising the step of configuring the object tracking program for setting at least one region of interest.

32. The method as claimed in claim 28 further comprising the step of configuring the object tracking program, said step comprises the steps of:

constructing an initial short term reference image and an initial long term reference image;

providing the object tracking program with the initial short term reference image and the initial long term reference image;

providing timing parameters; and assigning visual parameters.

33. The method as claimed in claim 32 wherein the step of constructing comprises creating the short term reference image and the long term reference image from a captured image.

34. The method as claimed in claim 32 wherein the step of constructing comprises creating the short term reference image and the long term reference image from internally stored images.

35. The method as claimed in claim 32 wherein the step of constructing comprises creating the short term reference image and the long term reference image through a learning process utilizing a set of sequentially ordered and captured images.

36. The method as claimed in claim 28 wherein the step pf pre-processing comprises the steps of:

obtaining the short term reference image;

obtaining the long term reference image;

obtaining a currently captured image;

generating a short term difference image from the short term reference frame and the currently captured image;

generating a long term difference image from the long term reference frame and the currently captured image.

37. The method as claimed in claim 28 wherein the step of clustering comprises the steps of:

building groups of clustered objects from at least two dynamic or static objects in accordance with the relative locations of each of the at least two dynamic or static objects;

adjusting the parameters of each of the at least two dynamic or static objects clustered within each group;

updating the parameters and status of each of the at least two dynamic or static objects.

38. The method of claim 28 wherein the step of clustering comprises the steps of predicting the motion of the at least one moving object by predictive filtering and adapting the parameters of the at least one moving object.

39. The method as claimed in claim 37 wherein the step of building groups of clustered objects comprises the steps of:

measuring the distance between each of the at least two dynamic or static objects;

determining neighborhood relations between each of the at least two dynamic or static objects and in accordance with the results of the distance measurement;

clustering the at least two dynamic or static objects in accordance with the determined neighborhood relations into distinct object groups; and

adjusting the distinct object groups in order to determine the optimal spatial parameters of each of the at least two dynamic or static objects in the distinct object groups.

40. The method as claimed in claim 38 wherein the step of adapting the parameters of the at least one moving object comprises the steps of locating the center of the at least one moving object; locating the boundary points constituting the boundary line of the at least one moving object; re-calculating the location of the center of the at least one moving object; and inserting the at least one moving object into an objects table.

41. The method as claimed in claim 40 further comprising the steps of adjusting the spatial parameters of the at least one moving object and retrieving similar objects to the at least one moving object from a discarded object archive.

42. The method as claimed in claim 29 wherein the step of characterizing comprises the steps of: measuring the movement of the at least one moving object to determine the behavior of the at least one moving object; merging spatially overlapping objects; generate an alarm trigger in accordance with the results of the behavior of the at least one moving object or in accordance with the spatial or visual parameters of the at least one moving object.

43. The method as claimed in claim 42 wherein the alarm trigger is generated in accordance with the texture of the object.

44. The method as claimed in claim 42 wherein the alarm trigger is generated in accordance with the shape of the object,

45. The method as claimed in claim 42 wherein the alarm trigger is generated in accordance with the velocity of the at least one moving object.

46. The method as claimed in claim 42 wherein the alarm trigger is generated in accordance with the trajectory of the at least one moving object.

47. The method as claimed in claim 28 wherein the step of updating the background comprises the steps of: updating the background draft;

updating the short term reference image; and updating the long term reference image.

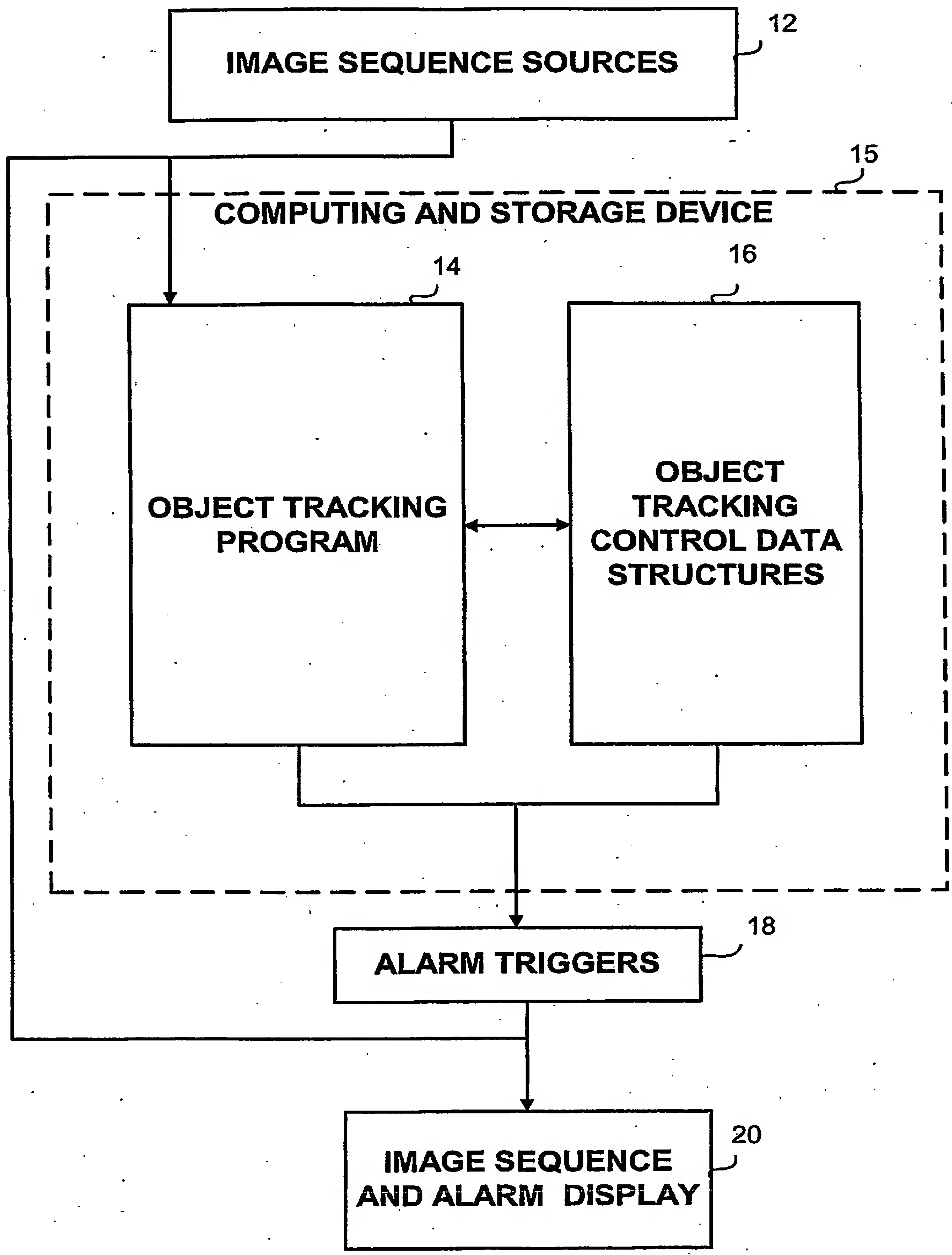


IMAGE SEQUENCE SOURCES
12
15
COMPUTING AND STORAGE DEVICE
14
16
OBJECT TRACKING PROGRAM
OBJECT TRACKING CONTROL DATA STRUCTURES
18
ALARM TRIGGERS
20
IMAGE SEQUENCE AND ALARM DISPLAY

FIG. 1

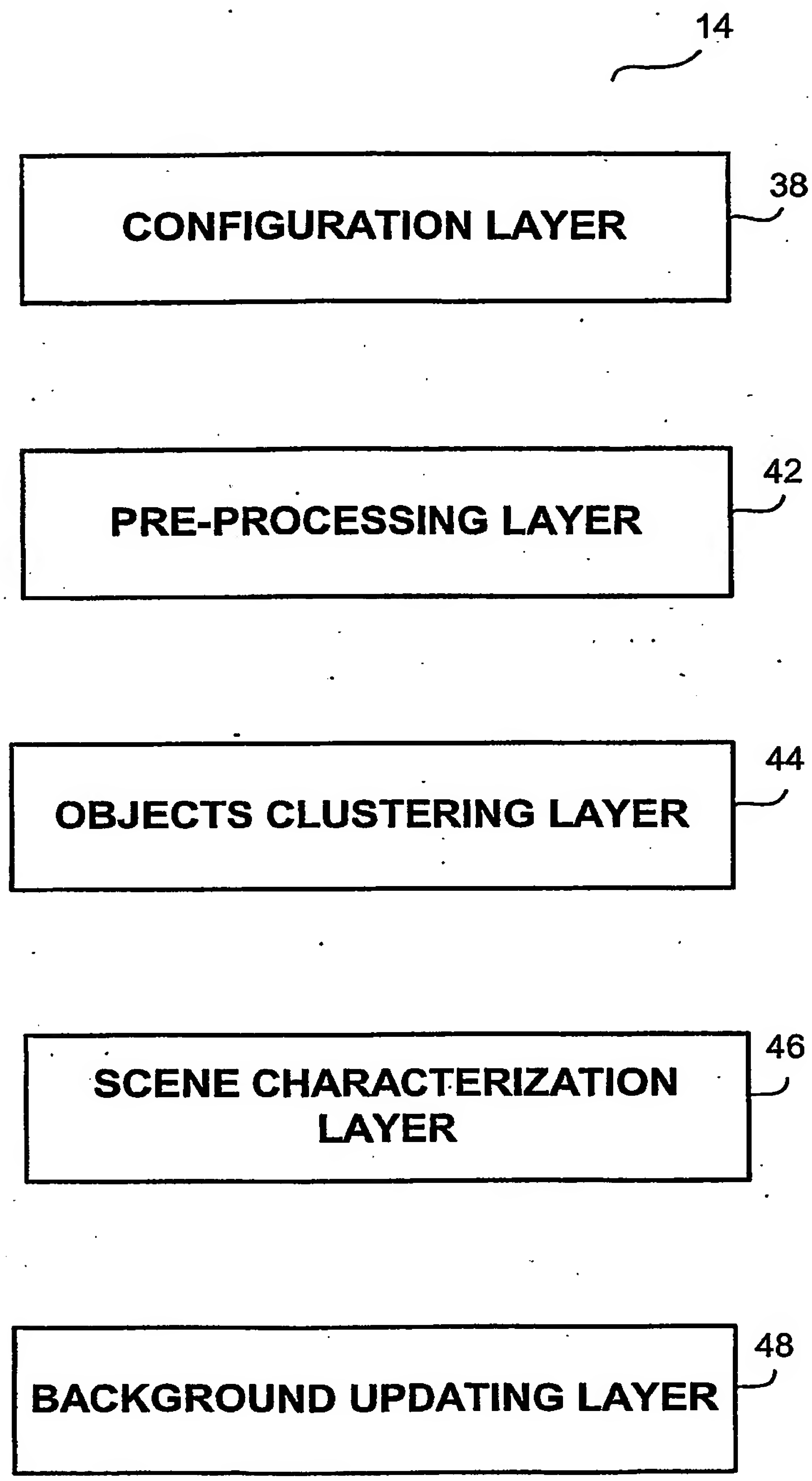

14
CONFIGURATION LAYER
38
PRE-PROCESSING LAYER
42
OBJECTS CLUSTERING LAYER
44
SCENE CHARACTERIZATION LAYER
46
BACKGROUND UPDATING LAYER
48

FIG. 2

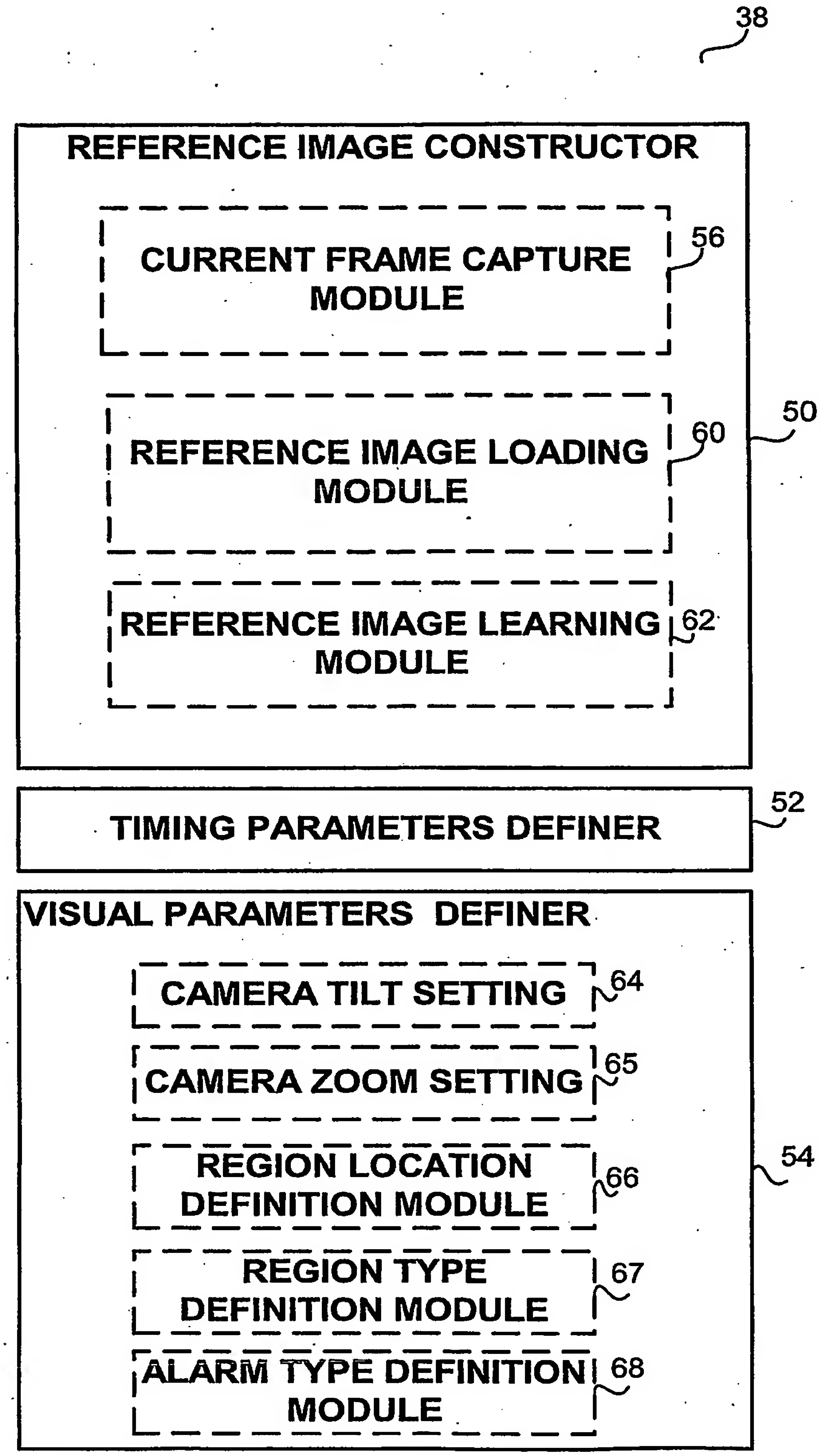

38
REFERENCE IMAGE CONSTRUCTOR
CURRENT FRAME CAPTURE MODULE
56
REFERENCE IMAGE LOADING MODULE
60
50
REFERENCE IMAGE LEARNING MODULE
62
TIMING PARAMETERS DEFINER
52
VISUAL PARAMETERS DEFINER
CAMERA TILT SETTING
64
CAMERA ZOOM SETTING
65
REGION LOCATION DEFINITION MODULE
66
54
REGION TYPE DEFINITION MODULE
67
ALARM TYPE DEFINITION MODULE
68

FIG. 3

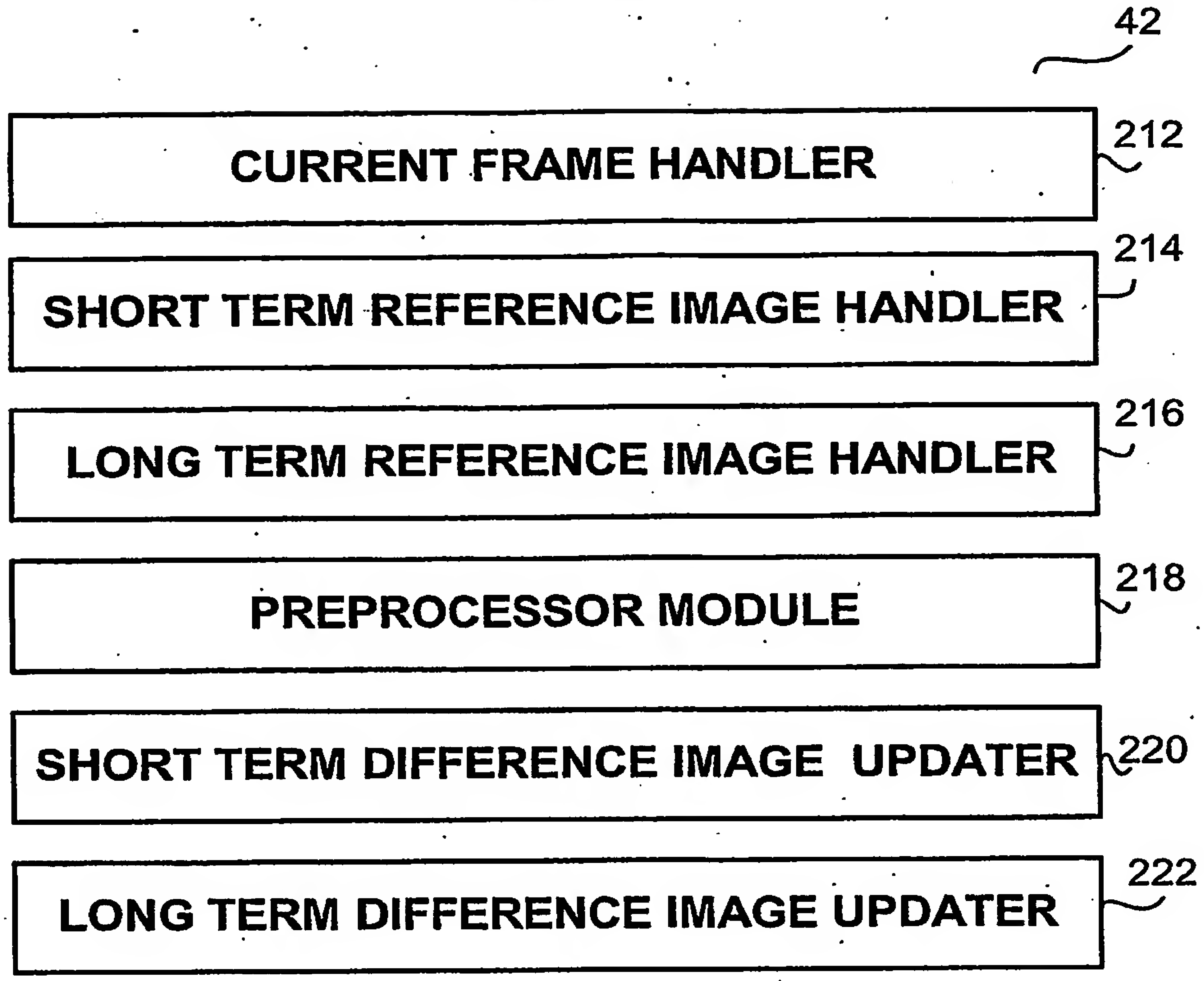

42
CURRENT FRAME HANDLER
212
SHORT TERM REFERENCE IMAGE HANDLER
214
LONG TERM REFERENCE IMAGE HANDLER
216
PREPROCESSOR MODULE
218
SHORT TERM DIFFERENCE IMAGE UPDATER
220
LONG TERM DIFFERENCE IMAGE UPDATER
222

FIG. 4A

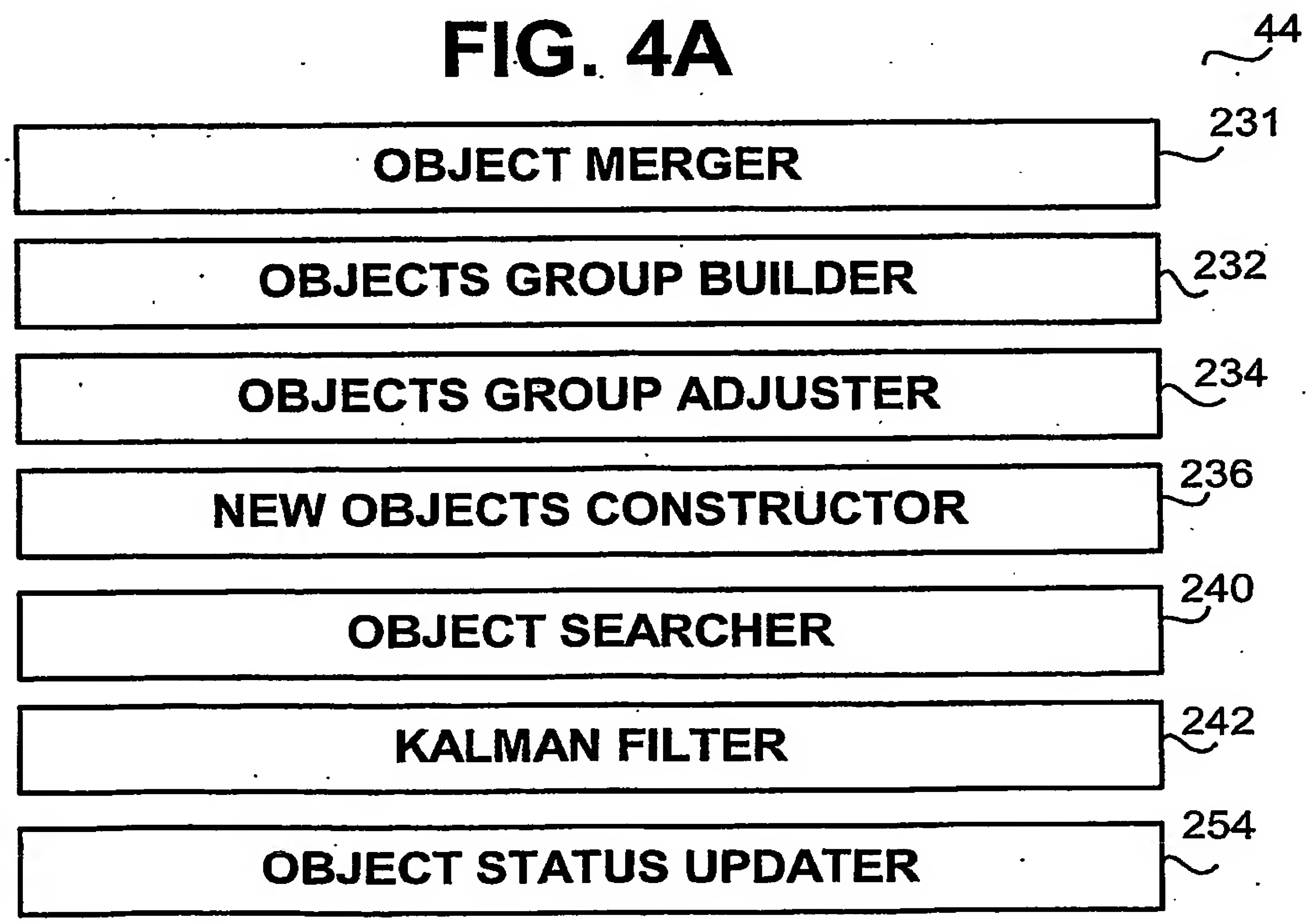

44
OBJECT MERGER
231
OBJECTS GROUP BUILDER
232
OBJECTS GROUP ADJUSTER
234
NEW OBJECTS CONSTRUCTOR
236
OBJECT SEARCHER
240
KALMAN FILTER
242
OBJECT STATUS UPDATER
254

FIG. 4B

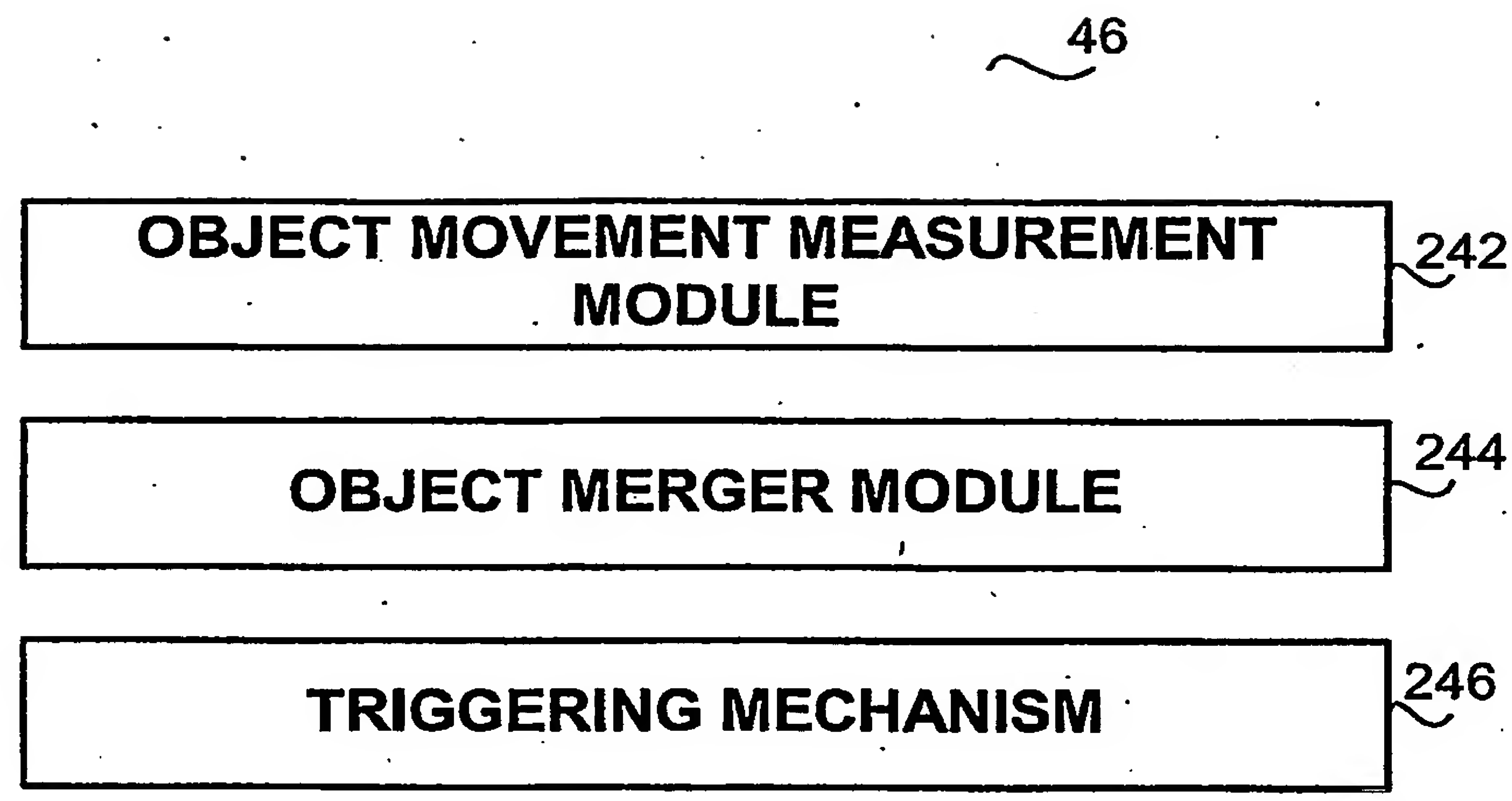

46
OBJECT MOVEMENT MEASUREMENT MODULE
242
OBJECT MERGER MODULE
244
TRIGGERING MECHANISM
246

FIG. 5A

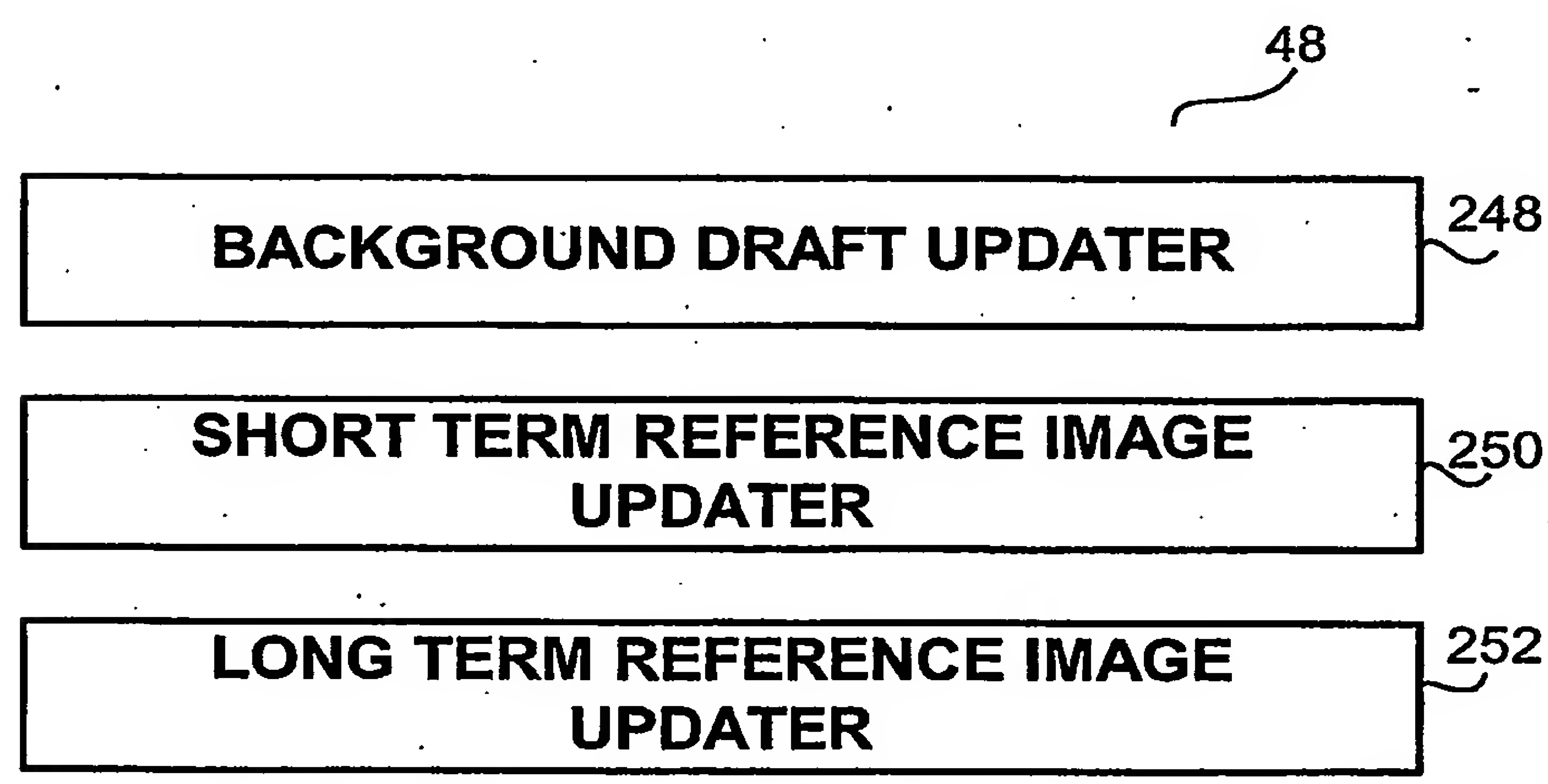

48
BACKGROUND DRAFT UPDATER
248
SHORT TERM REFERENCE IMAGE UPDATER
250
LONG TERM REFERENCE IMAGE UPDATER
252

FIG. 5B

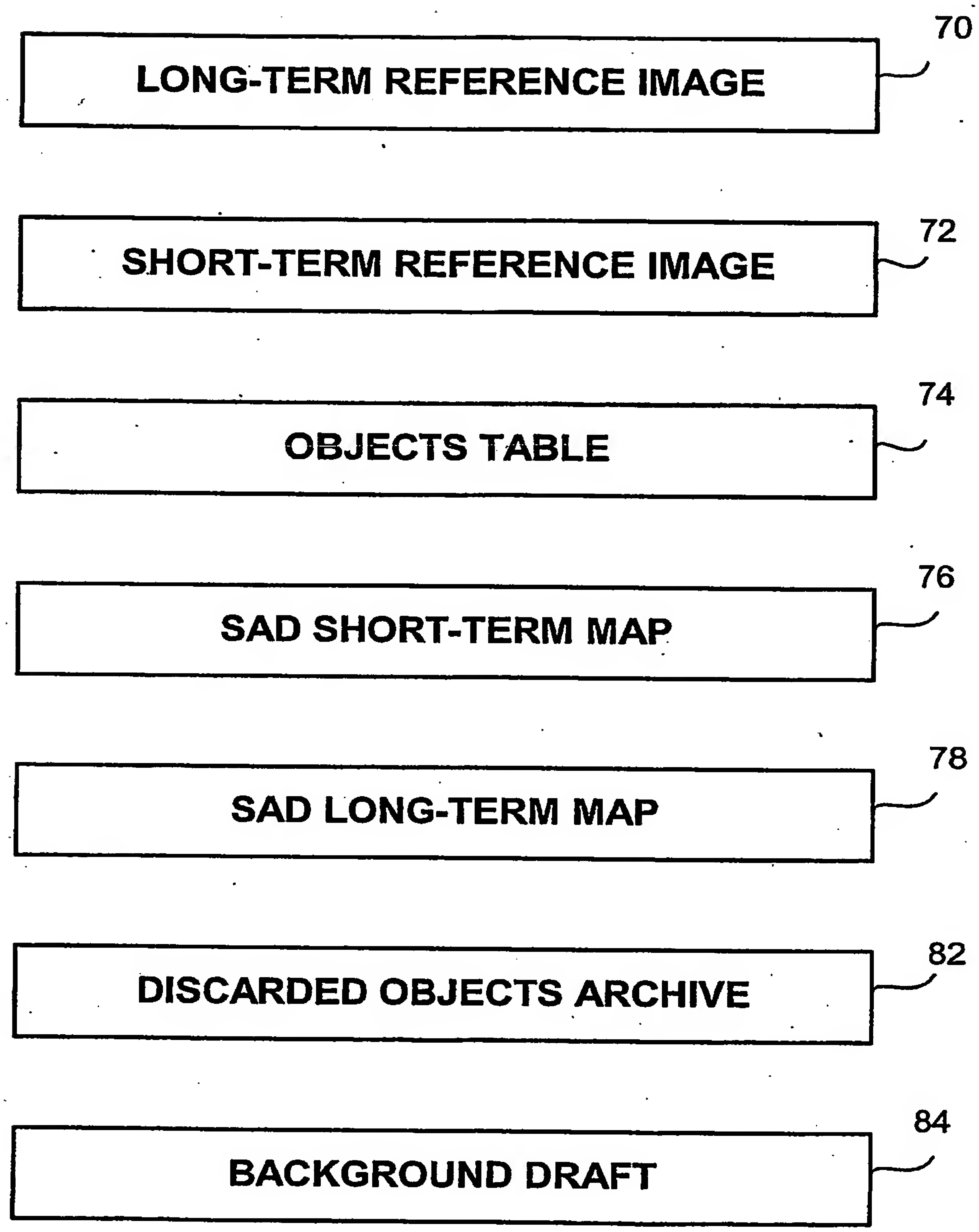

LONG-TERM REFERENCE IMAGE
70
SHORT-TERM REFERENCE IMAGE
72
OBJECTS TABLE
74
SAD SHORT-TERM MAP
76
SAD LONG-TERM MAP
78
DISCARDED OBJECTS ARCHIVE
82
BACKGROUND DRAFT
84

FIG. 6

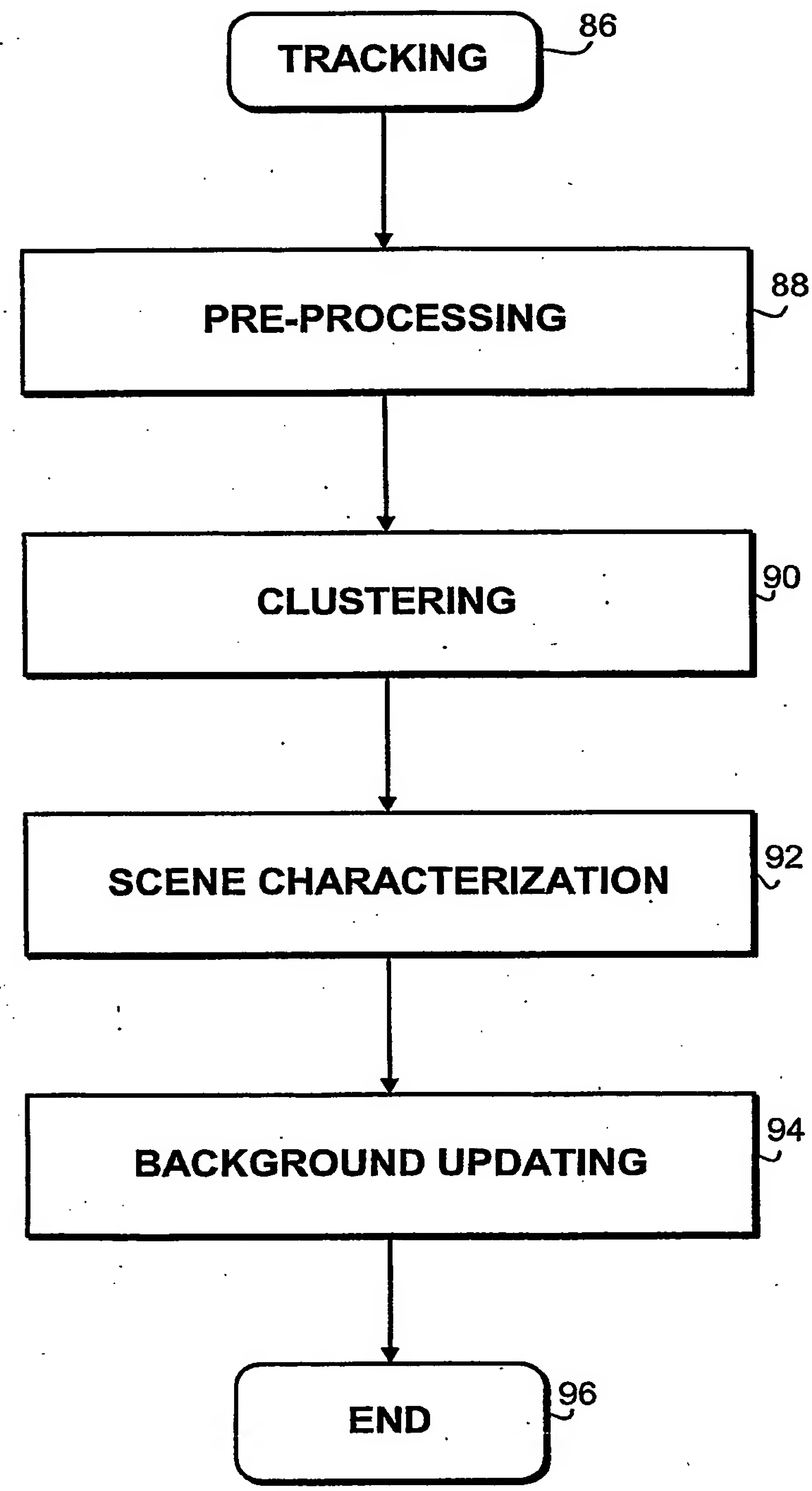

TRACKING
86
PRE-PROCESSING
88
CLUSTERING
90
SCENE CHARACTERIZATION
92
BACKGROUND UPDATING
94
END
96

FIG. 7

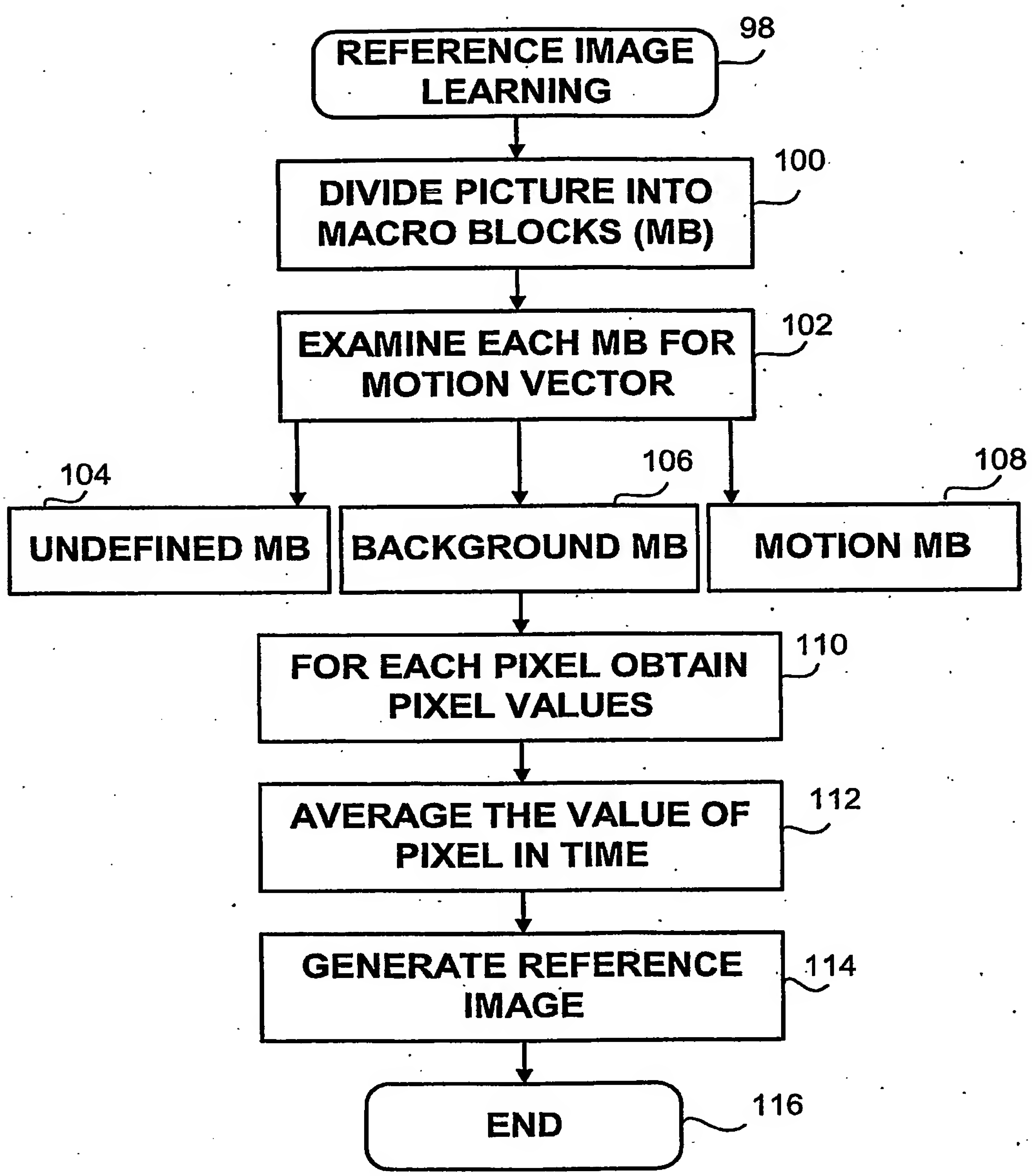

REFERENCE IMAGE LEARNING
98
DIVIDE PICTURE INTO MACRO BLOCKS (MB)
100
EXAMINE EACH MB FOR MOTION VECTOR
102
104
UNDEFINED MB
106
BACKGROUND MB
108
MOTION MB
FOR EACH PIXEL OBTAIN PIXEL VALUES
110
AVERAGE THE VALUE OF PIXEL IN TIME
112
GENERATE REFERENCE IMAGE
114
END
116

FIG. 8

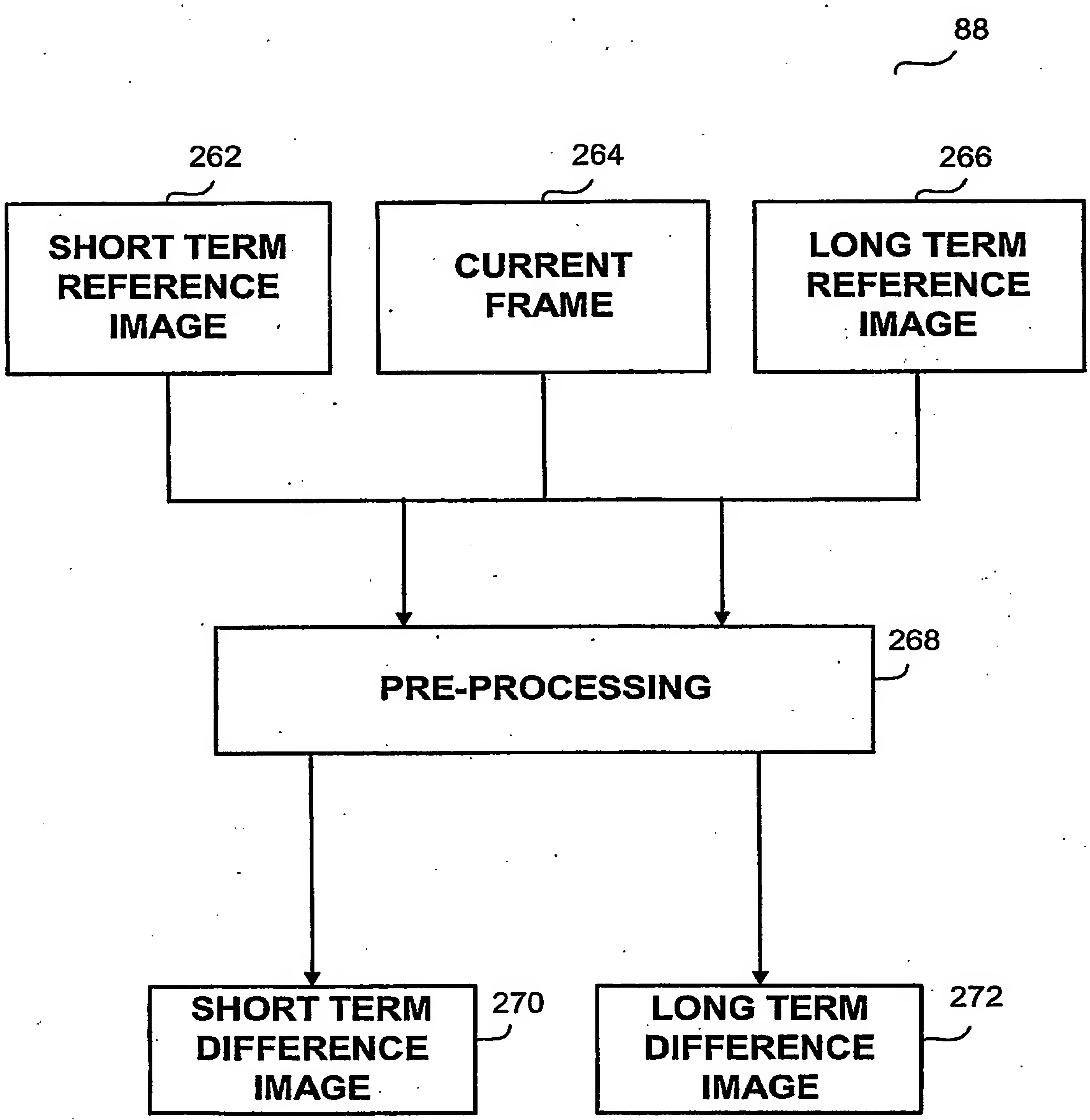

88
262
SHORT TERM REFERENCE IMAGE
264
CURRENT FRAME
266
LONG TERM REFERENCE IMAGE
PRE-PROCESSING
268
SHORT TERM DIFFERENCE IMAGE
270
LONG TERM DIFFERENCE IMAGE
272

FIG. 9

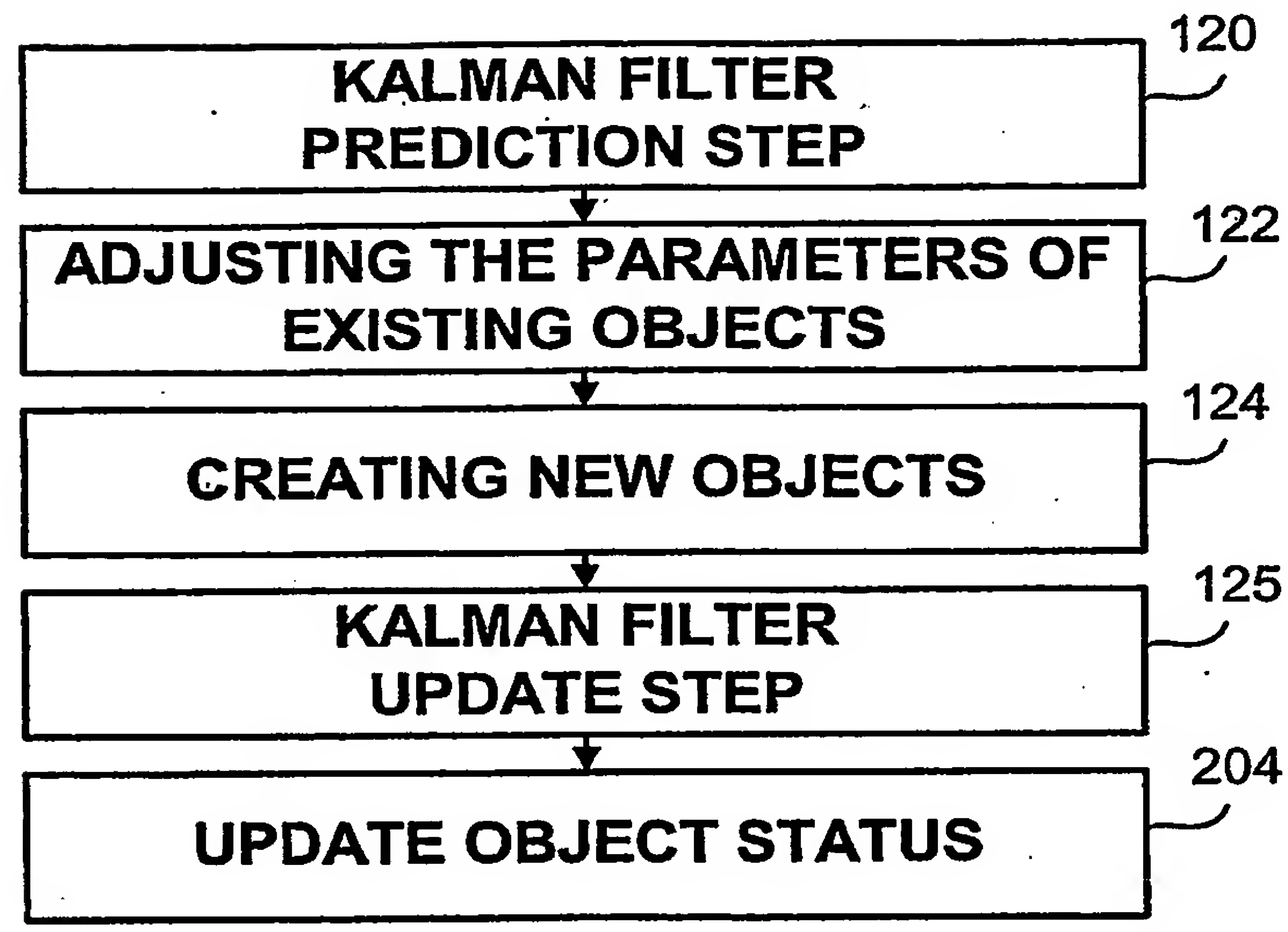


FIG. 10A

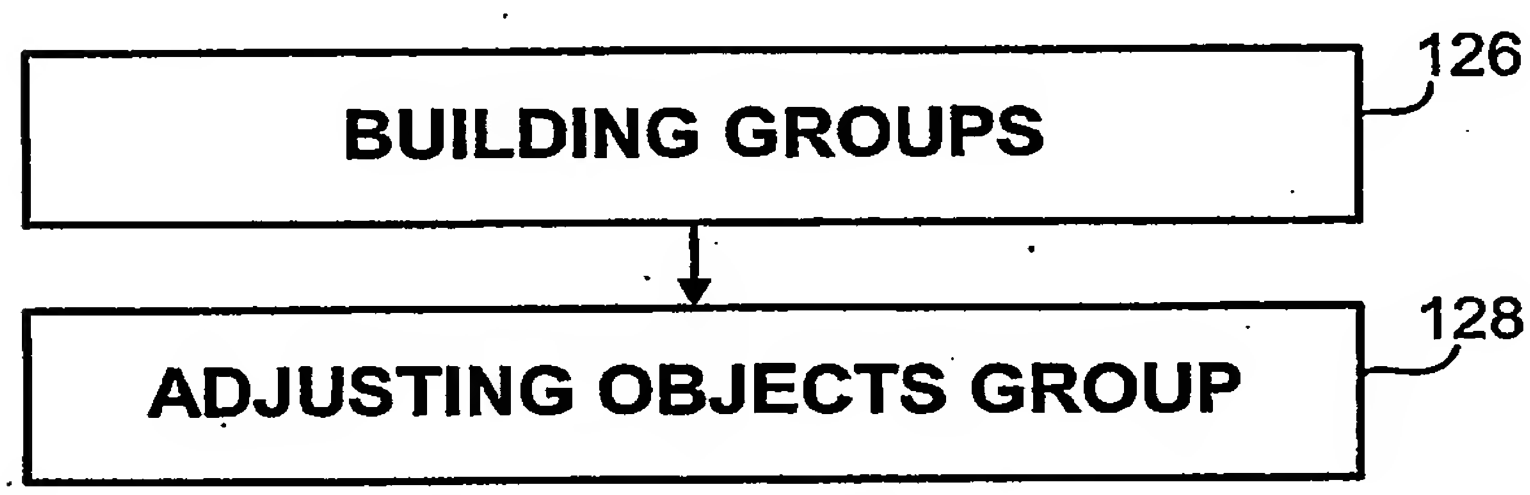


FIG. 10B

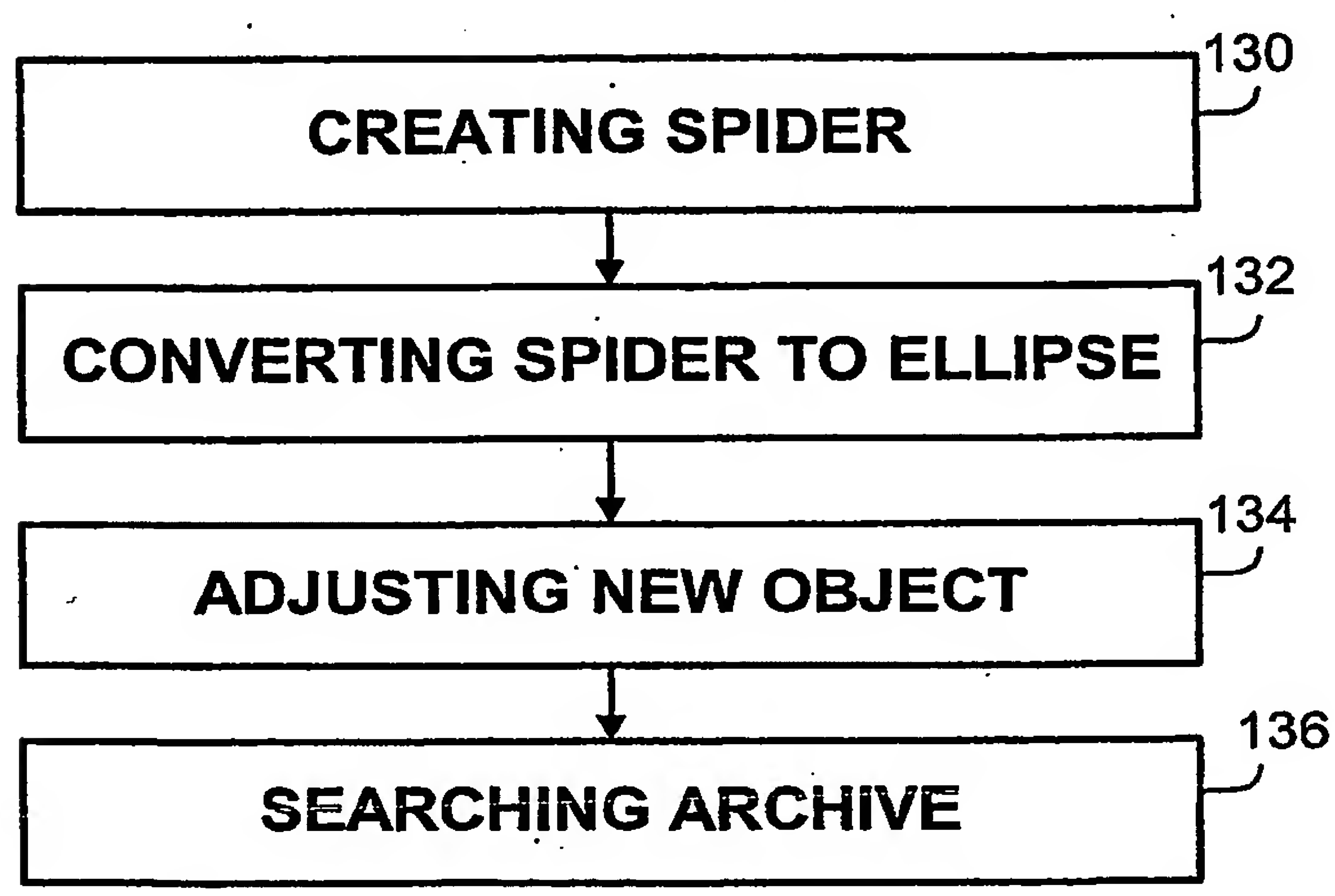


FIG. 10C

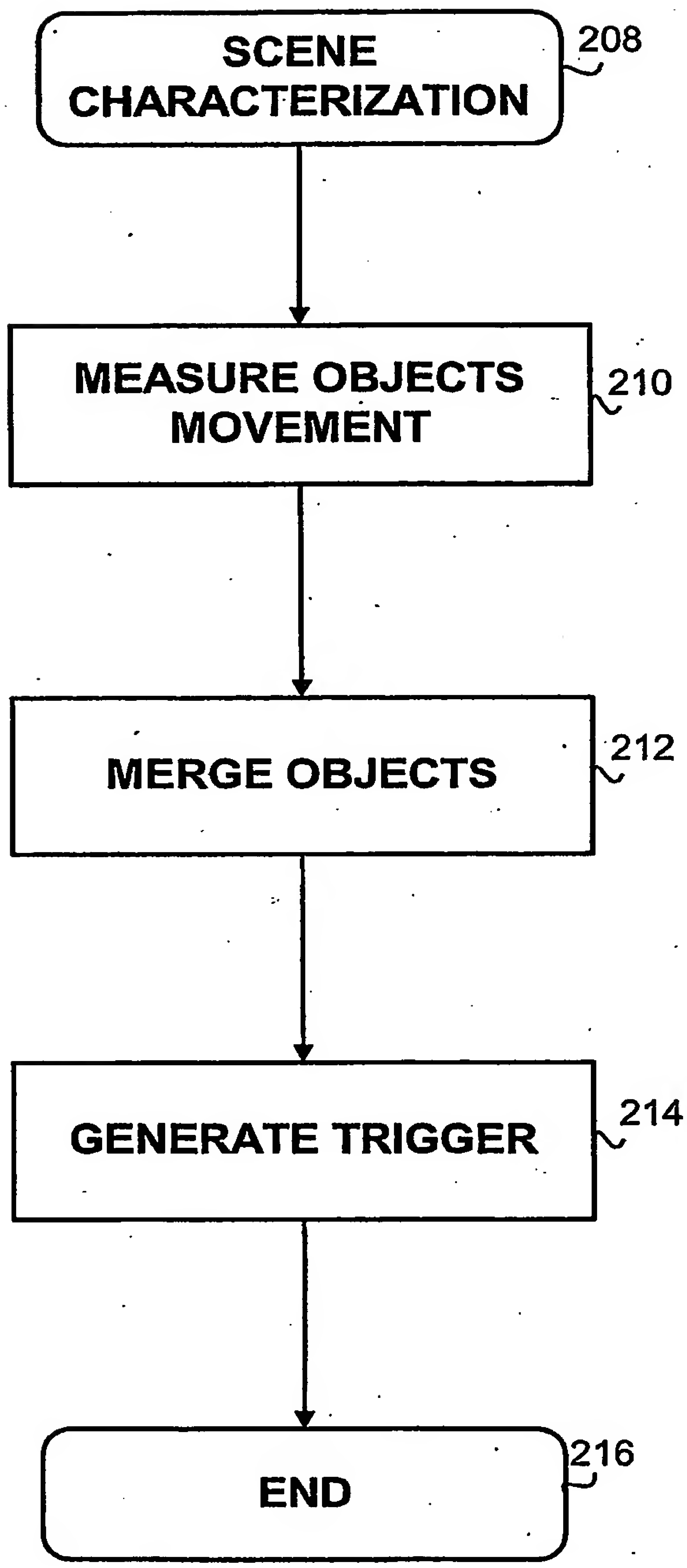

SCENE CHARACTERIZATION
208
MEASURE OBJECTS MOVEMENT
210
MERGE OBJECTS
212
GENERATE TRIGGER
214
END
216

FIG. 11

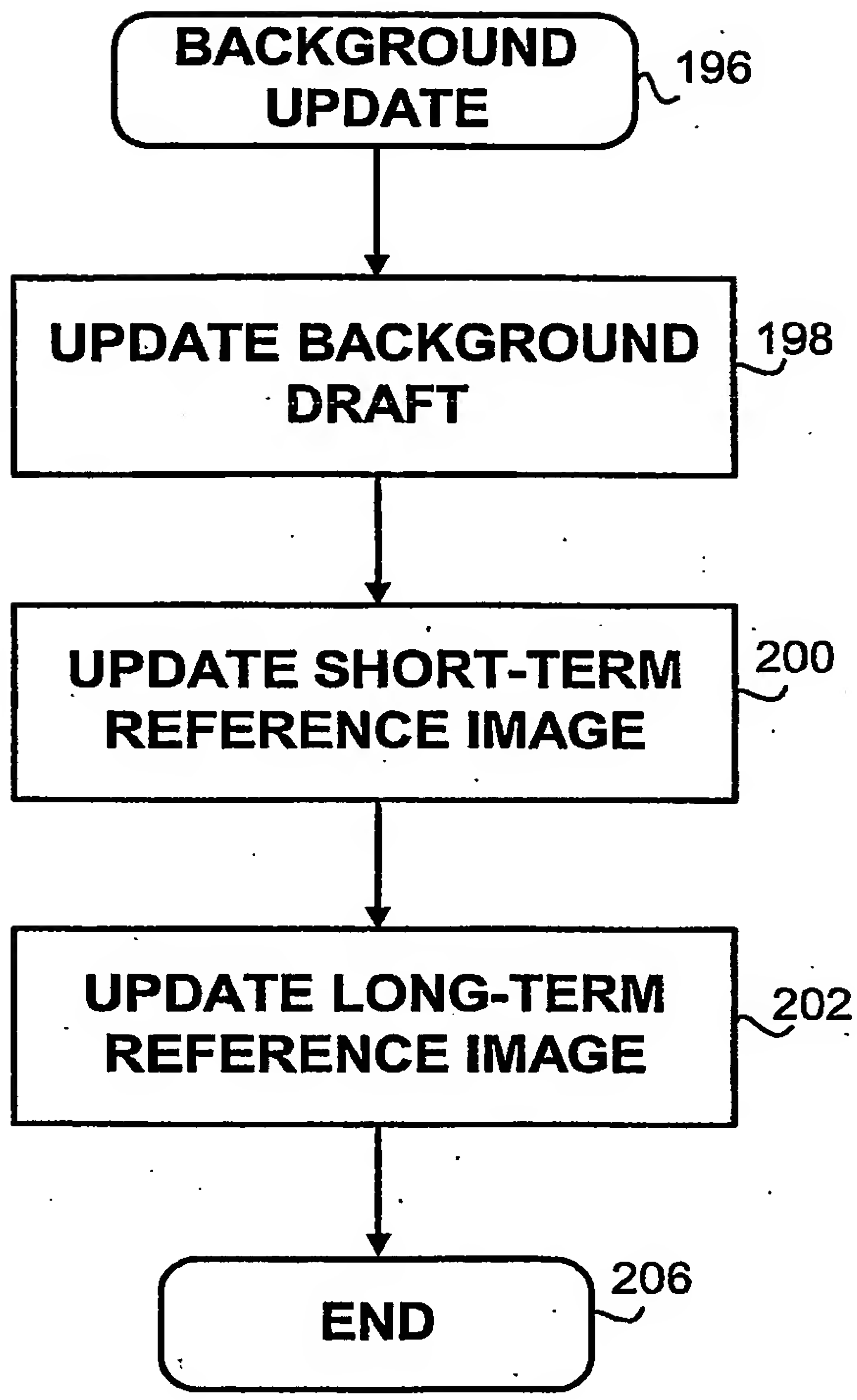

BACKGROUND UPDATE
196
UPDATE BACKGROUND DRAFT
198
UPDATE SHORT-TERM REFERENCE IMAGE
200
UPDATE LONG-TERM REFERENCE IMAGE
202
END
206

FIG. 12

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL03/00097

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : H04N 7/12, 5/225, 5/228
US CL : 375/240.08; 348/169, 208.14, 208.16
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 375/240.08; 348/169, 208.14, 208.16

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,091,780 A (POMERLEAU) 25 February 1992, figure 2, column 3, line 57 - column 6, line 21.	1-47
Y	US 5,307,170 A (ITSUMI et al) 26 April 1994, column 9, lines 25-40.	1-47
Y	US 5,734,441 A (KONDO et al) 31 March 1998 , figure 6, column8, lines 20-54.	1

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 May 2003 (19.05.2003)	05 JUN 2003

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Nhon T Diep Eugenia Logan
Facsimile No. (703)305-3230	Telephone No. 703 305-2600

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☑ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ______________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 October 2004 (21.10.2004)

PCT

(10) International Publication Number
WO 2004/091250 A1

(51) International Patent Classification[7]: H04Q 11/04, H04M 3/22

(21) International Application Number: PCT/EP2003/050098

(22) International Filing Date: 9 April 2003 (09.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (*for all designated States except US*): TELEFONAKTIEBOLAGET LM ERICSSON (publ) [SE/SE]; S-164 83 Stockholm (SE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): LAIHO, Keijo [FI/FI]; Metsätorpantie 2 G 20, FIN-02430 Masala (FI). VALASSI, Paolo [IT/IT]; Ericsson Telecommunicazioni SpA, Vimodrone/TEI/HD, Via Anagnina 203, CP 4235, I-00100 Roma Appio (IT).

(74) Agents: LIND, Robert et al.; Marks & Clerk, 4220 Nash Court, Oxford Business Park South, Oxford, Oxfordshire OX4 2RU (GB).

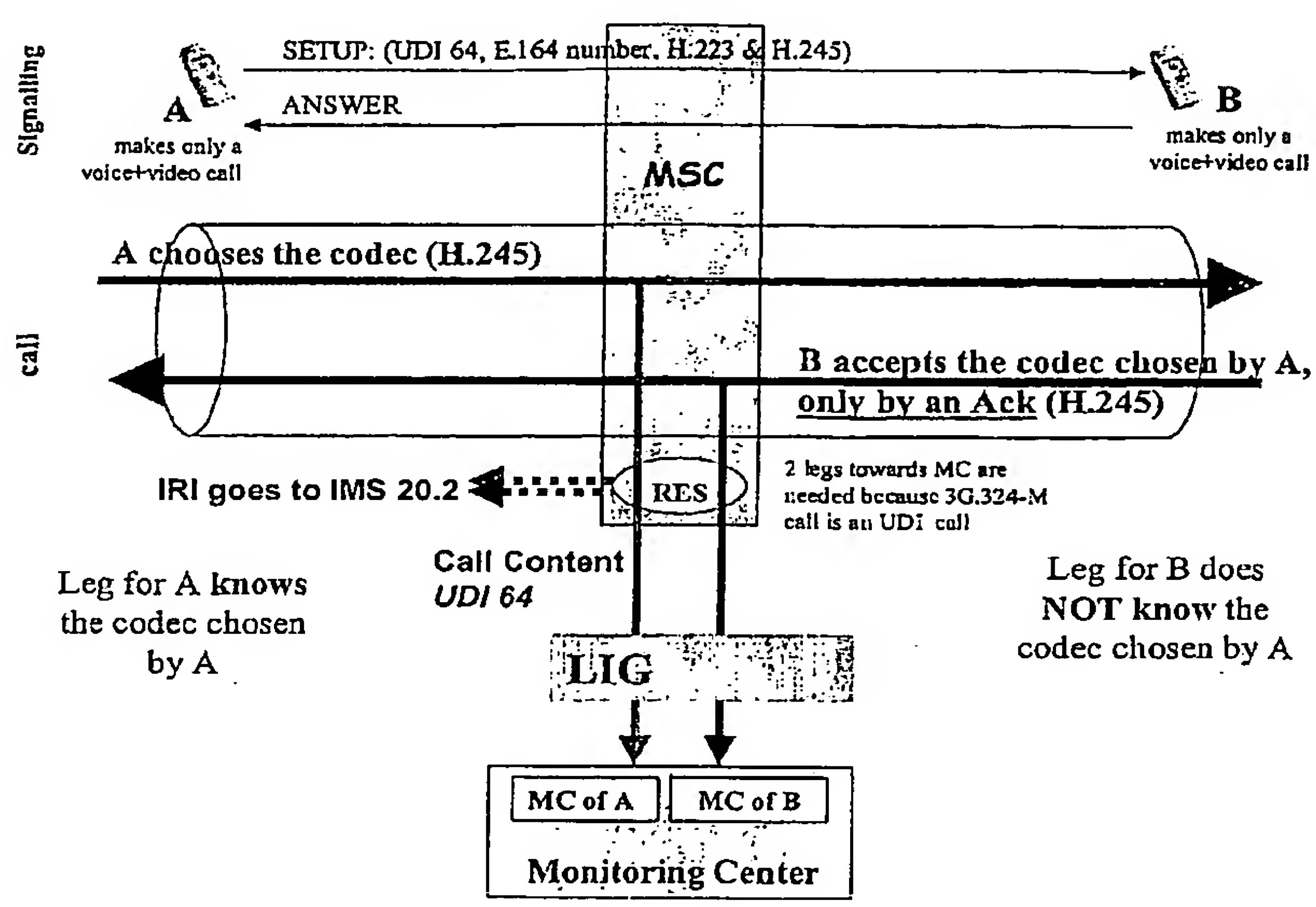
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: LAWFUL INTERCEPTION OF MULTIMEDIA CALLS



(57) **Abstract:** A method of performing lawful interception of a multimedia call between two or more terminals. The method comprises detecting the initiation of said call at monitoring equipment located in the call path, and forwarding from the monitoring equipment to a gateway, parameters defining at least one of the forward and reverse channels of said call. At least one multimedia call is set up from said gateway to a monitoring terminal in dependence upon the received parameters. Following the setting up of the first mentioned multimedia call, forward and/or reverse channel data is intercepted at said monitoring equipment, the intercepted data routed to said gateway, and the data transmitted to the monitoring terminal over the forward channel of the or each second mentioned multimedia call.

WO 2004/091250 A1

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Lawful Interception of Multimedia Calls

Field of the Invention

The present invention relates to the lawful interception of multimedia calls within a communications network.

Background to the Invention

The introduction of new communication systems including third generation mobile networks (3G) and broadband IP networks will result in a wide range of services being available to users. Not least amongst these services will be the possibility for multimedia (MM) calls between users, allowing video telephony and the exchange of data.

There are circumstances in which authorised agencies such as the police and intelligence services must be able to monitor calls including multimedia calls. Such lawful interception is required in order to be able to collect information on those suspected of involvement in criminal or terrorist activities. The lawful interception of traditional voice call has been handled in two ways:

1) The voice streams coming from the subscribers involved in a call to be intercepted are mixed together by monitoring equipment located in one of the "switches" involved in the call. The mixed stream is sent, by establishing an ancillary call, to the monitoring centre. Thus the mixed stream, i.e. the complete conversation between the parties, can be played for example using an ordinary loudspeaker in the monitoring centre.

2) The voice streams coming from the subscribers involved in the intercepted call are not mixed, but rather two connections are established from the monitoring equipment to the monitoring centre, each carrying one leg of the call. This allows the monitoring centre to record the voices of the two call parties separately and/or mix the voice streams in the monitoring centre.

The lawful interception of multimedia calls is more problematic than for voice calls. The protocols used to set up a multimedia call between terminals require handshaking between the participating terminals. The handshaking is used to agree upon parameters describing the payload of the call and how the payload is to be transported. The parameters to describe the payload include a used codec and codec options (e.g. video codecs such as H263 and MPEG4 include a number of optional features, the main purpose of which are to either improve the

picture quality or decrease the used bandwidth, or both). Transport parameters include for example payload format, e.g. the format of the RTP-packet to be used to carry a data stream in IP based transport network, or H223 logical channel parameters used in narrowband multimedia H.324. H223 logical channel parameters include parameters specifying whether payload frames are allowed to be segmented into several H223 multiplex frames, whether the payload frames are numbered, etc.

Figure 1 illustrates for example a handshake between two terminals according to the ITU-T H.245 protocol (where "OLC" designates Open Logical Channel signaling messages). In the lawful interception scenario, it is not possible to involve the monitoring centre in the handshaking process as two terminals are already involved in the process and in any case it is undesirable to alert a terminal associated with a call to the interception action. For multimedia calls therefore, according to current interception processes, normal multimedia equipment (e.g. mobile handsets) cannot be used in the monitoring centre to decode and display the media. Interception can only be achieved using specialist equipment installed at the monitoring centre.

Summary of the Invention

According to a first aspect of the present invention there is provided a method of performing lawful interception of a multimedia call between two or more terminals, the method comprising:

detecting the initiation of said call at monitoring equipment located in the call path;

forwarding from the monitoring equipment to a gateway, parameters defining at least one of the forward and reverse channels of said call;

setting up at least one multimedia call from said gateway to a monitoring terminal in dependence upon the received parameters; and

following the setting up of the first mentioned multimedia call, intercepting forward and/or reverse channel data at said monitoring equipment, routing the intercepted data to said gateway, and transmitting the data to the monitoring terminal over the forward channel of the or each second mentioned multimedia call.

A main function of the gateway is to map, where necessary, protocols used in the network connecting the terminals involved in the call being intercepted, to protocols used in the network connecting the gateway to the monitoring terminal. These protocols include media control protocols (e.g. H.245), call control protocols (ISUP, H.225), multiplexing protocols (H.223), and audio and video codec protocols.

In one embodiment of the present invention, said terminals are H.324 terminals and a multimedia call is established between these terminals via circuit switched networks. The monitoring terminal is an H.323 or SIP terminal, and communicates with said gateway via a broadband IP network.

Preferably, said monitoring equipment forwards to said gateway, signalling messages exchanged between the terminals involved in the call being intercepted. The gateway uses the information contained in these messages to setup the multimedia call(s) to the monitoring terminal and/or to setup transcoding functions within the gateway. The need for transcoding is determined primarily by the properties of the monitoring terminal, as well as the properties of the gateway.

The method may comprise setting up a call from said gateway to the monitoring terminal for each of the forward and reverse channels of the intercepted call. Alternatively, the forward and reverse channels data may be multiplexed/mixed onto the forward channel of a single call established between said gateway and the monitoring terminal. In another alternative, two calls may be established between the gateway and respective terminals at the monitoring centre. Forward channel data from the intercepted call is placed on the forward channel of one of these two calls, whilst reverse channel data is placed on the forward channel of the other one of the calls.

According to a second aspect of the present invention there is provided apparatus for intercepting a multimedia call between two or more terminals, the apparatus comprising:

means for receiving from monitoring equipment located within the call path, parameters defining at least one of the forward and reverse channels of said call, following detection of the initiation of said call by the monitoring equipment;

means for setting up at least one multimedia call to a monitoring terminal; and

means for receiving intercepted forward and/or reverse channel data from said monitoring equipment, and for transmitting the data to a monitoring terminal over the forward channel(s) of the second mentioned multimedia call(s).

Brief Description of the Drawings

Figure 1 illustrates handshake signalling between two H.324 terminals;

Figure 2 illustrates schematically a Video Interactive Gateway providing an interface between H.324 and H.323 domains;

Figure 3 illustrates the use a Lawful Interception Gateway to intercept two calls between H.324 terminals;

Figure 4 shows in detail signalling between two H.324 terminals, and between a Lawful Interception Gateway and an H.323 monitoring terminal;

Figure 5 shows signalling between two SIP terminals, and between a Lawful Interception Gateway and a SIP monitoring terminal; and

Figure 6 illustrates network nodes involved in lawful interception where calls are set up using SIP.

Detailed Description of a Preferred Embodiment

The following standards will be referred to *inter alia* in this description of a preferred embodiment of the present invention:

ITU-T H.323	*Packet based multimedia communications systems;*
ITU-T H.324	*Terminal for low bit-rate multimedia communication;*
ITU-T H.223	*Multiplex protocol for low bit rate multimedia communication;*
ITU-T H.245	*Control protocol for multimedia communication;*
3GPP TS 24.228	*Signalling flows for the IP multimedia call control based on SIP and SDP;*
3GPP TS 33.108	*Handover interface for lawful interception.*

By way of explanation, there is now provided a general outline of the various protocols used to establish and control multimedia calls, and of the protocols defining multimedia data types. There will then be provided a description of an embodiment of the invention which provides for the lawful interception of multimedia calls.

Multimedia calls can be divided into two categories: multimedia calls using narrowband circuit connections and multimedia calls using an IP (broadband) network.

In the case of multimedia calls transported over narrowband circuit connections, a known protocol is ITU-T H.324. H.324 uses a mechanism in which different multimedia components are multiplexed into a single data stream, which is transported over the circuit connection. H.223 is used by H.324 as a multiplexing protocol, to multiplex different data streams from

different media codecs (e.g. G.723, AMR for audio, and H.263, MPEG4 for video) and the media control protocol (H.245) into a single data stream. The circuit switched call itself might typically be established using ISUP.

In the case of multimedia calls transported over an IP-network, known protocols in this category for establishing and controlling calls are H.323 and Session Initiation Protocol (SIP). The fundamental mechanism for these two protocols is the same. The media control protocol is transported via a TCP/IP (or SCTP/IP) connection between terminals. The media streams are transported by using separate RTP/IP connections for each media between the terminals. H.323 uses H.225 to set up connections between H.323 terminals.

Interworking between these two categories of multimedia calls is generally achieved by using a so-called Video Interactive Gateway (VIG) which makes possible interworking between low bit-rate multimedia terminals (H.324) located in circuit switched telephony networks and terminals in IP based multimedia systems (H.323/SIP). The circuit switched networks may use the 64 kbit/s unrestricted digital bearer for the multimedia connection. Using H.223 as the multiplexing protocol, different multimedia components (audio, video and data) are multiplexed within the circuit switched bearer. These channels are de-multiplexed by the VIG onto separate RTP and TCP channels in the IP network, and *vice versa*. VIG may perform transcoding for different multimedia components if necessary in order to make communication between end terminals possible.

H.245 may be used as a control protocol both in circuit switched networks and in IP networks, providing end-to-end capability exchange, signalling of command and indications, and messages to open and describe the content of logical channels for different multimedia components. The VIG performs mapping of H.245 messages between a circuit switched network and an IP network, in order to adapt the different transport protocols and to enable transcoding of media channels. The VIG will perform mapping if necessary between the call control protocol in the circuit switched network (ISUP), and that in the IP network (H.225).

Figure 2 illustrates schematically a VIG interfacing H.324 and H.323 networks. The VIG comprises a Media Gateway operating at the bearer level and providing interworking between user data, and a Media Gateway Controller operating at the call control level and providing interworking between signalling protocols.

It must be possible to carry out the lawful interception of calls between terminals regardless of the protocols used between the terminals. However, this should be possible using some standard piece of equipment on the part of the intercepting authority, i.e. it is not desirable to have to select equipment depending upon the protocols used between callers and upon whether or not a VIG is present in a call path.

Figure 3 shows an example lawful interception scenario for a call between two mobile terminals (A and B) having narrowband access (e.g. via a 3G network), both terminals being H.324 terminals. Monitoring equipment (this essentially being equipment for placing a "tap" on both legs of a call) is located within an MSC of the GSM network. A Lawful Interception Gateway (LIG) provides a gateway between the monitoring equipment and a monitoring centre. The monitoring centre comprises an H.323 terminal coupled to the LIG via a broadband IP network. In a typical scenario, the H.323 terminal at the monitoring centre is implemented on a standard Personal Computer (PC). Whilst the PC might use, for example, Microsoft Netmeeting™ to establish calls with the LIG, the LIG would typically use a proprietary solution for this purpose.

The LIG acts as a VIG (see Figure 2), translating data between the narrowband and broadband formats. The functions performed by the LIG are as follows:

- The LIG listens to the incoming data streams from the monitoring equipment.
- It decodes the transport/multiplex protocols (e.g. H.223)
- The LIG decodes the relevant information from the media control protocol, i.e. codec information within the Session Description Protocol (SDP) in case where SIP is used in the broadband network, and codec information and other information (e.g. H.223 logical channel parameters within H.245 in case of H.323).
- The LIG establishes a connection to a normal multimedia terminal in the monitoring centre based on the received information.
- The LIG emulates a normal multimedia terminal towards the normal multimedia terminal within the monitoring centre, by performing the complete media control protocol transactions with that terminal. This includes: 1) invoking the required procedures to connect the media streams for the data coming from the monitoring equipment, and 2) responding correctly to the procedure invocations coming from the terminal in the monitoring centre.
- The LIG forwards the media streams coming from the monitoring equipment, over the established connections to the monitoring centre.

Figure 4 illustrates signalling exchanges between the H.324 terminals A and B. In order to set up a call between the two terminals, a terminal capabilities exchange procedure (or handshake) is performed. The results of this negotiation are confirmed by terminal A to terminal B in an OLC (Forward Channel Description, Reverse Channel Description) message. The MSC in which the monitoring equipment is located maintains or has access to a database of subscribers for whom lawful interception warrants have been served. When a MM call is initiated to or from a subscriber on whom such an order has been placed, the MSC notifies the LIG. The MSC then forwards to the LIG the entire (64kbit/s) multiplexed streams, in both the forward and reverse directions, including the OLC (Forward Channel Description, Reverse Channel Description) message sent from terminal A to terminal B.

The LIG examines the parameters of the two legs of the call, and initiates two calls to the H.323 terminal at the monitoring centre. The properties of the forward channel (i.e. which will carry data from the LIG to the monitoring centre) of the first call correspond to the properties of the forward channel of the call between terminals A and B. The properties of the forward channel of the second call correspond to the properties of the reverse channel of the call between terminals A and B. The properties of the reverse channels of the two calls between the LIG and the H.323 terminal are irrelevant as these channels will not be used to carry "live" data.

An assumption here is that the H.323 terminal at the monitoring centre is able to terminate two calls simultaneously, and therefore that the forward and reverse channels of the intercepted call can be carried on respective calls to that H.323 terminal. An alternative mechanism is for the LIG to establish calls to two different H.323 terminals at the monitoring centre, or for a single call to be established with the forward and reverse channel data being multiplexed/mixed onto that single call. An appropriate mechanism may be selected by the LIG based upon a terminal capabilities negotiation with the H.323 terminal.

The LIG may include transcoding capabilities, which makes it possible to use multimedia terminals in the monitoring centre which do not support all possible codecs.

Figure 5 illustrates signaling in a scenario where the terminal used at the monitoring centre utilises SIP signaling to establish calls over a broadband IP network to which the LIG is also attached, and in which the two terminal participating in the intercepted call also use SIP signaling. Again, following notification of (forward and reverse channel) parameters by the monitoring equipment at the MSC, the LIG establishes two calls to the SIP terminal at the

monitoring centre. It will be appreciated that in this embodiment of the invention the LIG does not provide any VIG functionality.

Figure 6 illustrates in more detail the interception procedure. Within the IP multimedia subsystem (IMS), a Proxy CSCF (P-CSCF) participates in SIP signalling. The P-CSCF may be located either in a participating terminal's home network or in a visited network to which that terminal is attached. The P-CSCF identifies the SIP-URL(s) to which SIP signalling belongs. The P-CSCF also has a knowledge of SIP-URLs for which calls are to be intercepted. Using this information, the P-CSCF forwards SIP signalling associated with a call to be intercepted to the LIG as shown in Figure 5 (the LIG is implemented as part of the Delivery Function (DF)). The P-CSCF commands the GPRS Gateway Support Node (GGSN) to make a copy of RTP-stream (media streams) and forward them it to the LIG. In Figure 6, the monitoring terminal corresponds to the LEMF node, the latter being 3GPP terminology. According to 3GPP, the H3 and H2 interfaces carry user and signalling data respectively from the interception node to the monitoring terminal. According to the present invention, these interfaces are "merged" into one or more multimedia calls.

It will be appreciated by the person of skill in the art that various modifications may be made to the above described embodiment. For example, the LI subscriber database available to the MSC may define for subscribers on whom an interception warrant has been place, whether the reverse and forward channels are to be intercepted, or whether only one of these channels is to be intercepted. This information is signaled to the LIG.

Whilst in the scenario described with reference to Figure 3 the terminals A and B are H.324 terminals whilst the intercepting terminal is an H.323 terminal, other scenarios are possible. These include:

1. A and B terminals are H.324 terminals. The monitoring centre has an H323 terminal. The LIG performs H245-H245 mapping between two half calls and two complete calls. The LIG also performs TDM/H223 to IP/RTP interworking.
2. A and B terminals are H324 terminals. The monitoring centre has an H324 terminal. The LIG performs H245-H245 mapping between two half calls and two complete calls.

3. A and B terminals are H324 terminals. The monitoring centre has a SIP terminal. The LIG performs H245-SIP mapping between two half calls and two complete calls. The LIG also performs TDM/H223 to IP/RTP interworking.
4. A and B terminals are SIP terminals. The monitoring centre has an H323 terminal. The LIG performs SIP-H245 mapping between two half calls and two complete calls.
5. A and B terminals are SIP terminals. The monitoring centre has an H324 terminal. The LIG performs SIP-H245 mapping between two half calls and two complete calls. The LIG also performs TDM/H223 to IP/RTP interworking.
6. A and B terminals are SIP terminals. The monitoring centre has a SIP terminal. The LIG performs SIP-SIP mapping between two half calls and two complete calls.

In the SIP embodiment of Figure 5, it might sometimes be the case that intercepted data does not need conversion/transcoding at the LI gateway. In that case, the P-CSCF might instruct the GGSN to forward intercepted data directly to the monitoring terminal. No multimedia call need be established between the LI gateway and the monitoring terminal.

Claims

1. A method of performing lawful interception of a multimedia call between two or more terminals, the method comprising:

detecting the initiation of said call at monitoring equipment located in the call path;

forwarding from the monitoring equipment to a gateway, parameters defining at least one of the forward and reverse channels of said call;

setting up at least one multimedia call from said gateway to a monitoring terminal in dependence upon the received parameters; and

following the setting up of the first mentioned multimedia call, intercepting forward and/or reverse channel data at said monitoring equipment, routing the intercepted data to said gateway, and transmitting the data to the monitoring terminal over the forward channel of the or each second mentioned multimedia call.

2. A method according to claim 1, said gateway performing a mapping between protocols used in the network connecting the terminals involved in the call being intercepted, to protocols used in the network connecting the gateway to the monitoring terminal.

3. A method according to claim 1 or 2, wherein the monitoring terminal communicates with said gateway via a broadband IP network.

4. A method according to any one of the preceding claims, said monitoring equipment forwarding to said gateway, signalling messages exchanged between the terminals involved in the call being intercepted.

5. A method according to any one of the preceding claims, said gateway performing transcoding of intercepted channel data.

6. A method according to any one of the preceding claims and comprising setting up a call from said gateway to the monitoring terminal for each of the forward and reverse channels of the intercepted call.

7. A method according to any one of claims 1 to 5 and comprising multiplexing/mixing the intercepted forward and reverse channel data onto the forward channel of a single call established between said gateway and the monitoring terminal.

8. A method according to any one of claims 1 to 5 and comprising establishing two calls between the gateway and respective terminals at the monitoring centre, forward channel data from the intercepted call being placed on the forward channel of one of these two calls, whilst reverse channel data is placed on the forward channel of the other one of the calls.

9. A method according to any one of the preceding claims, wherein the terminals participating in the first mentioned multimedia call are H.324 terminals, and said monitoring terminal is an H.323 terminal.

10. A method according to any one of claims 1 to 8, wherein the terminals participating in the first mentioned multimedia call are SIP terminals, and said monitoring terminal is also a SIP terminal

11. Apparatus for intercepting a multimedia call between two or more terminals, the apparatus comprising:

means for receiving from monitoring equipment located within the call path, parameters defining at least one of the forward and reverse channels of said call, following detection of the initiation of said call by the monitoring equipment;

means for setting up at least one multimedia call to a monitoring terminal; and

means for receiving intercepted forward and/or reverse channel data from said monitoring equipment, and for transmitting the data to a monitoring terminal over the forward channel(s) of the second mentioned multimedia call(s).

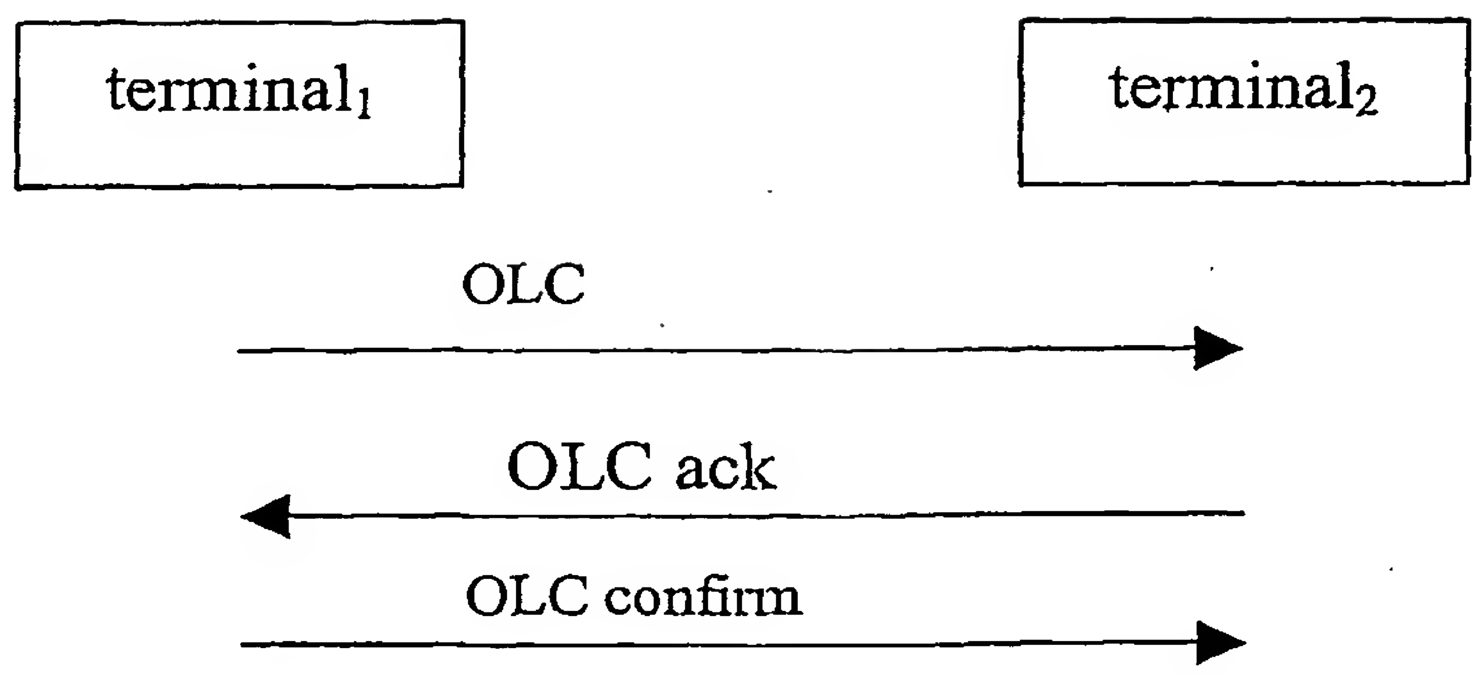


Figure 1

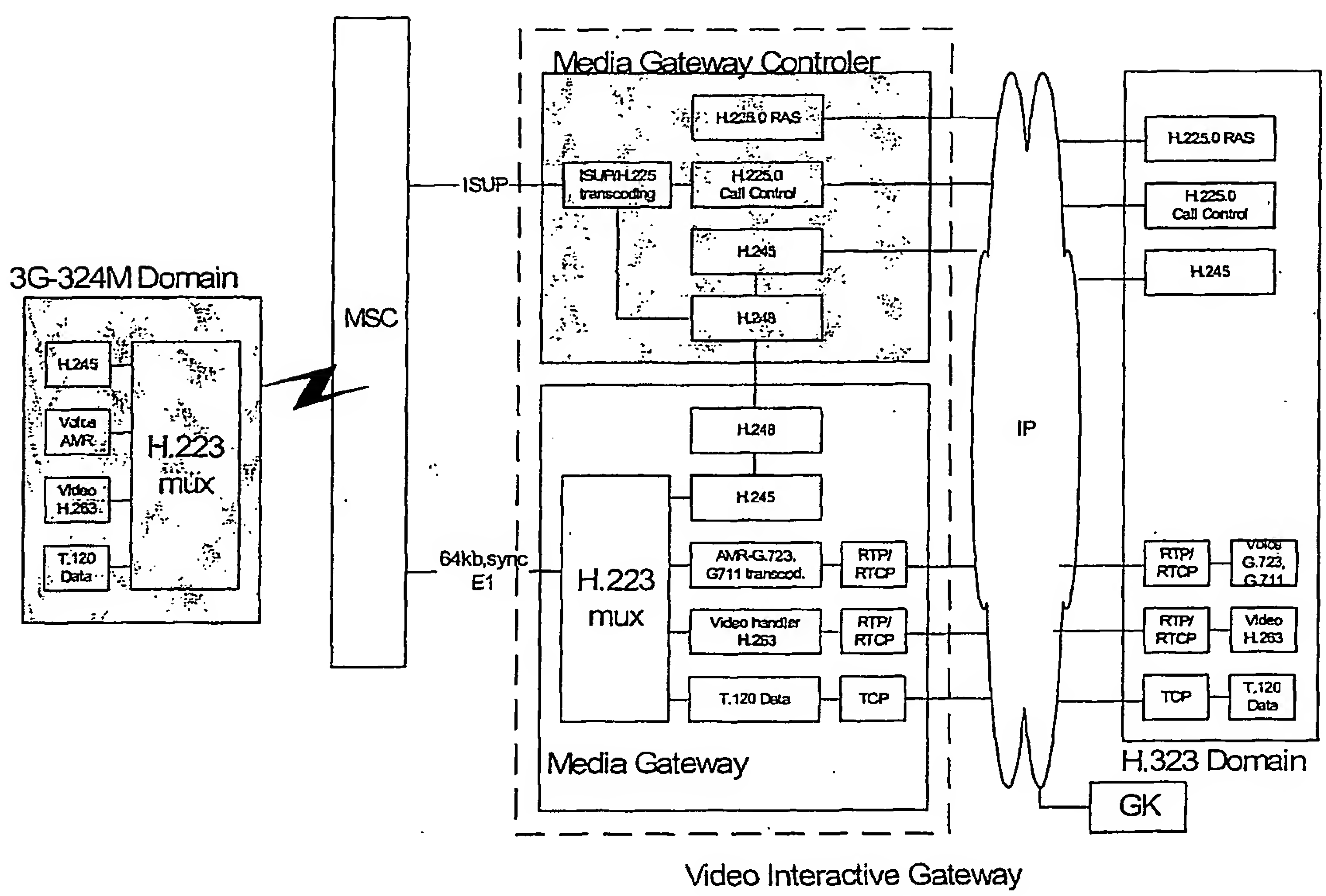


Figure 2

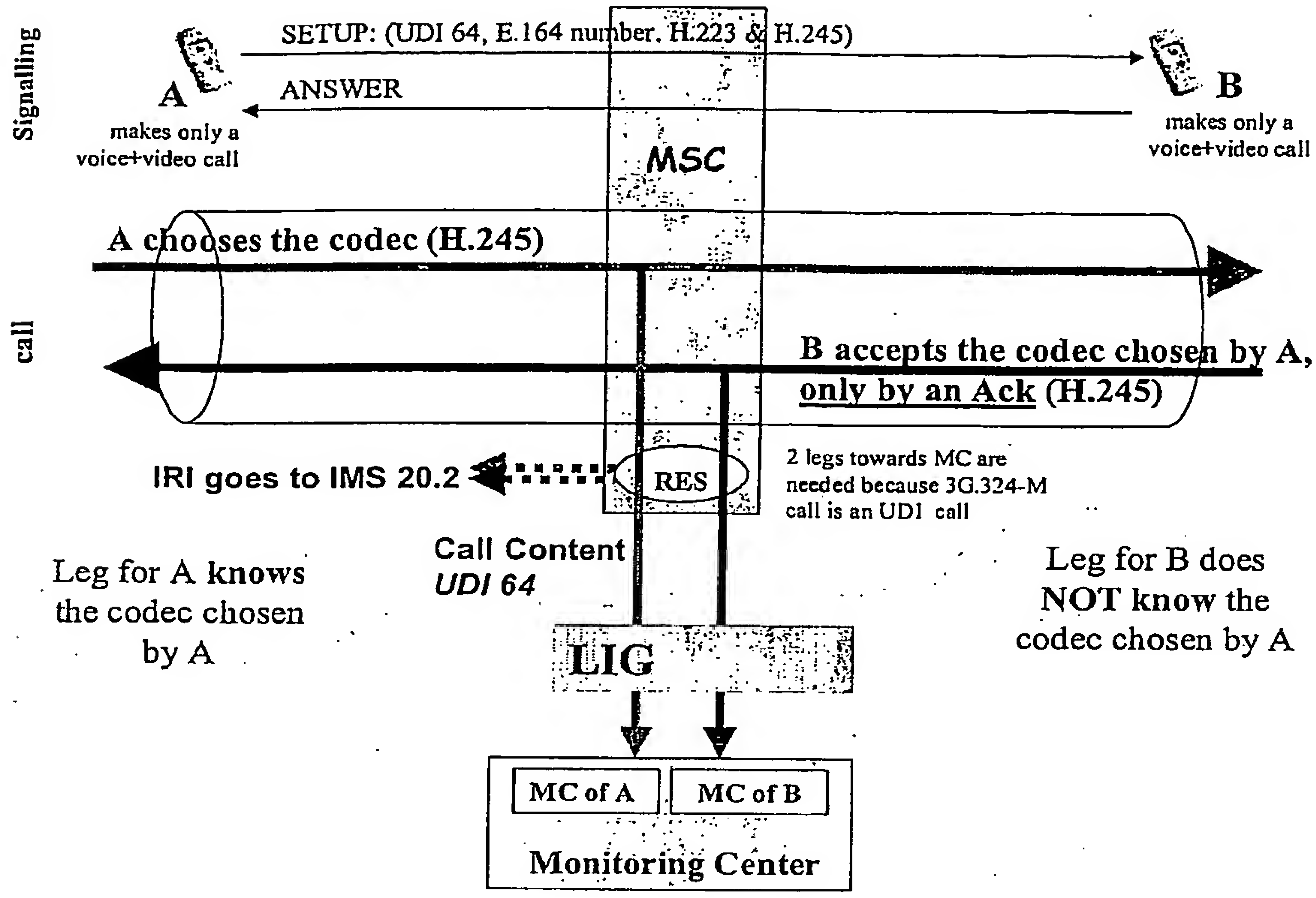


Figure 3

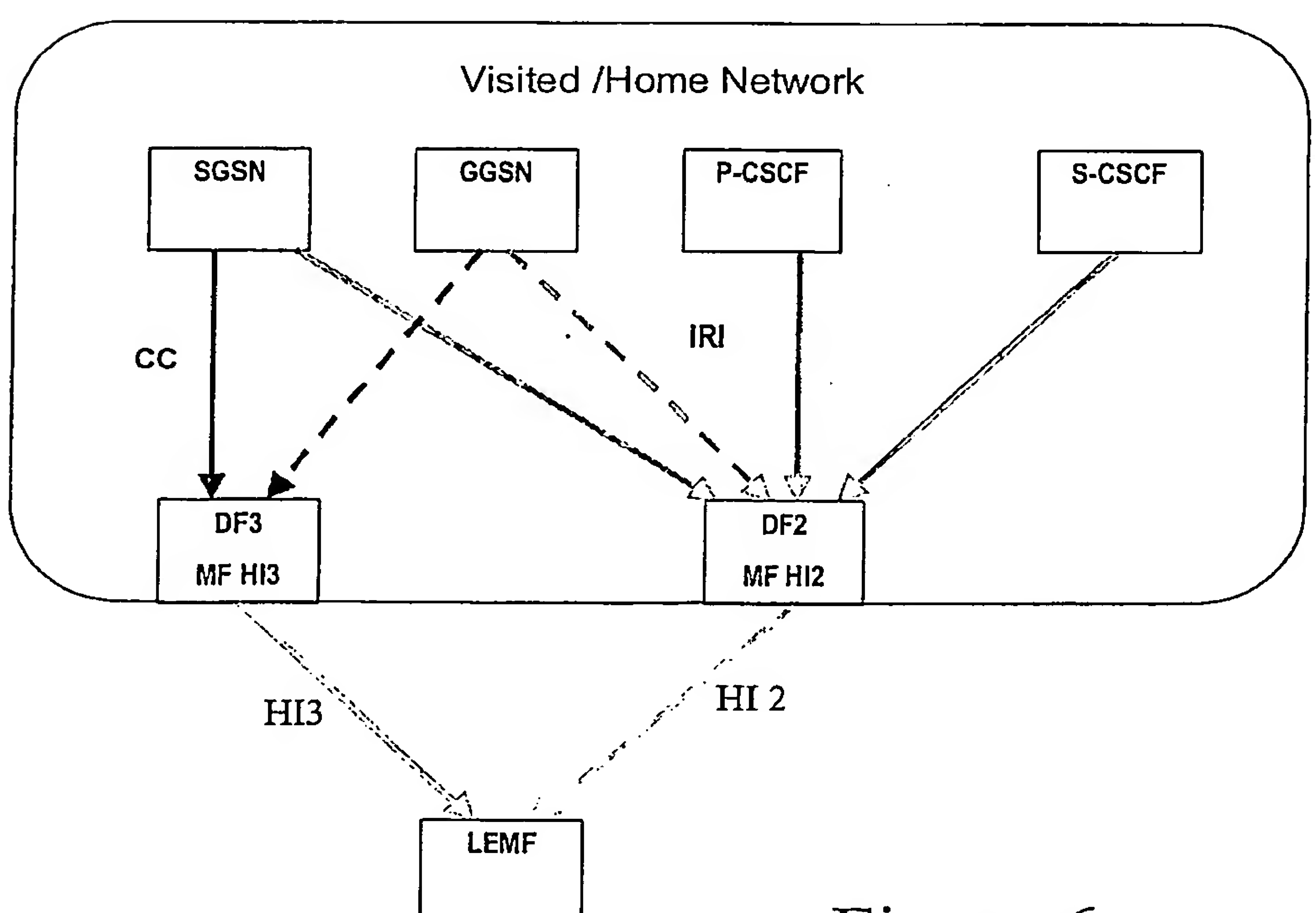


Figure 6

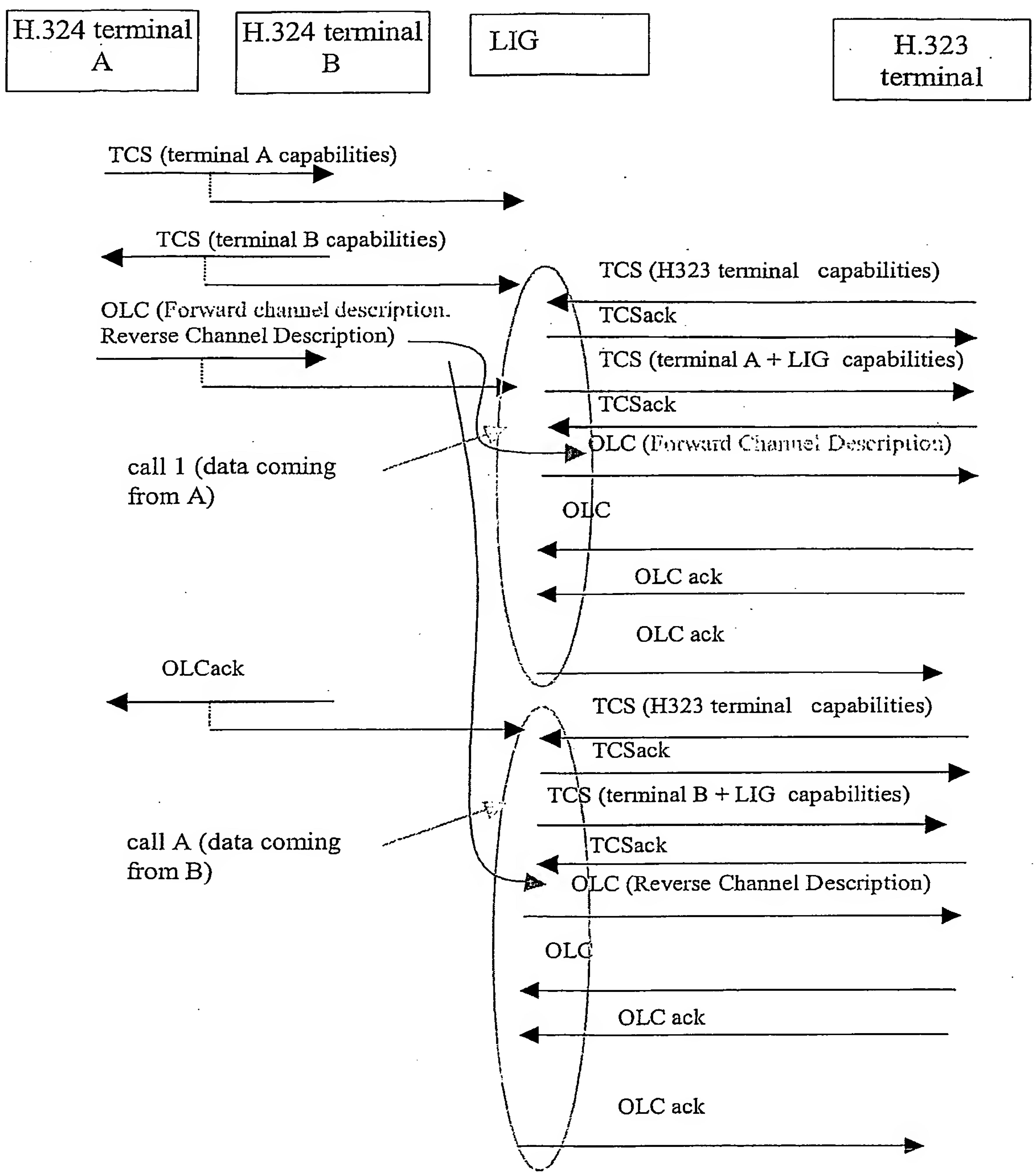

H.324 terminal A
H.324 terminal B
LIG
H.323 terminal
TCS (terminal A capabilities)
TCS (terminal B capabilities)
TCS (H323 terminal capabilities)
TCSack
OLC (Forward channel description. Reverse Channel Description)
TCS (terminal A + LIG capabilities)
TCSack
OLC (Forward Channel Description)
call 1 (data coming from A)
OLC
OLC ack
OLC ack
OLCack
TCS (H323 terminal capabilities)
TCSack
TCS (terminal B + LIG capabilities)
call A (data coming from B)
TCSack
OLC (Reverse Channel Description)
OLC
OLC ack
OLC ack

Figure 4

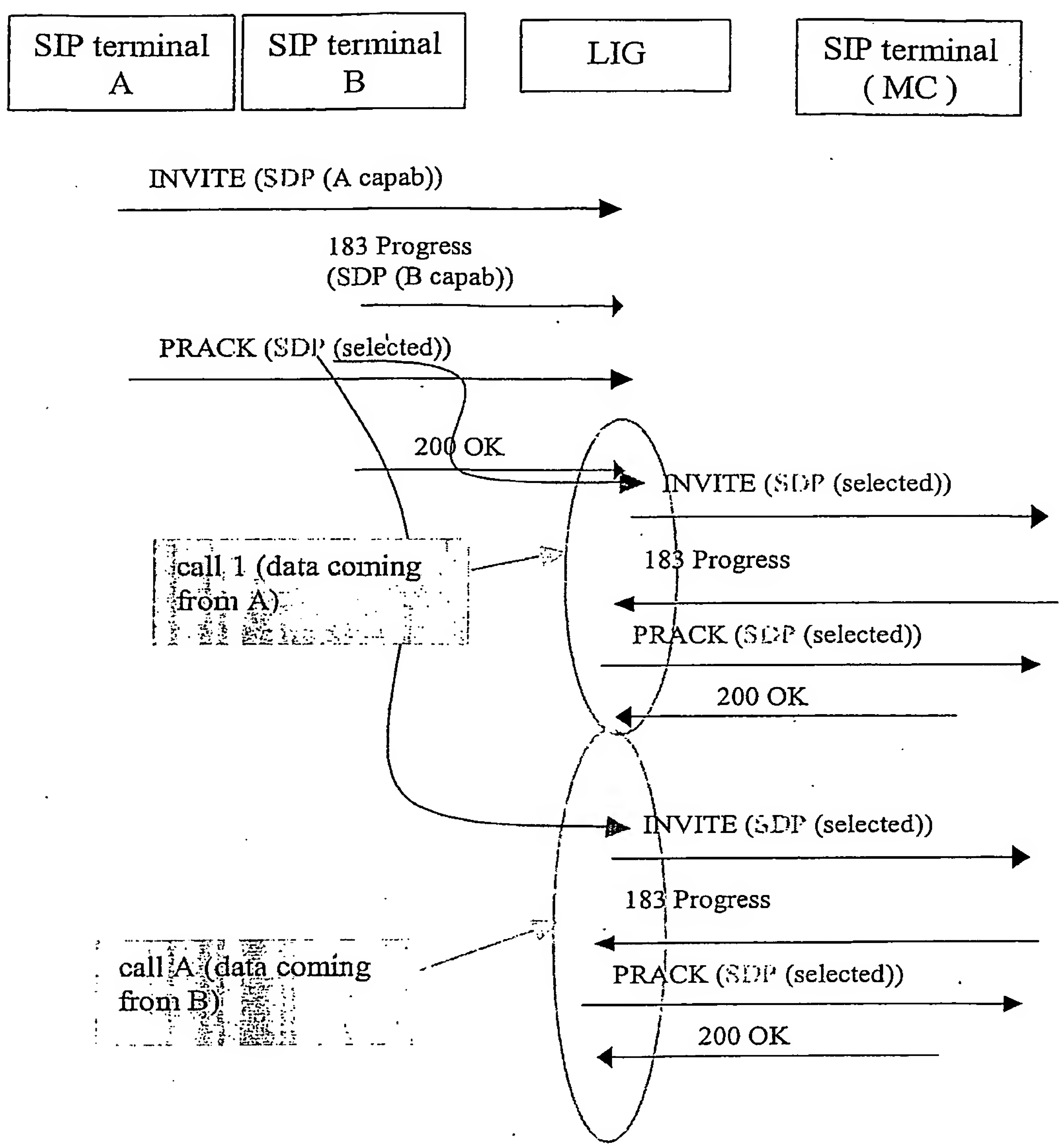

SIP terminal A
SIP terminal B
LIG
SIP terminal (MC)
INVITE (SDP (A capab))
183 Progress (SDP (B capab))
PRACK (SDP (selected))
200 OK
INVITE (SDP (selected))
call 1 (data coming from A)
183 Progress
PRACK (SDP (selected))
200 OK
INVITE (SDP (selected))
183 Progress
call A (data coming from B)
PRACK (SDP (selected))
200 OK

Figure 5

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

International Appl[redacted]n No
PCT/EP 03/50098

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04Q11/04 H04M3/22

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04Q H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, IBM-TDB, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 02 093838 A (NOKIA CORP) 21 November 2002 (2002-11-21) page 10, line 7 -page 12, line 31 claims 1-18,23-42; figure 1	1-4,6-11 5
X,L	WO 01 91374 A (TELEFON AB LM ERICSSON) 29 November 2001 (2001-11-29) abstract page 2, line 23 -page 3, line 27 page 13, line 10 -page 16, line 18 claims 1,13; figure 1	1-4,11

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 February 2004	04/03/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Gijsels, W

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

International Appli[redacted]n No
PCT/EP 03/50098

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02 085041 A (T-MOBILE DEUTSCHLAND GMBH) 24 October 2002 (2002-10-24) abstract page 9, line 1-24 page 13, line 15-27 claims 1,3,5,7,11-21; figures 1,3	1-4,11
X Y	WO 01 47222 A (ERICSSON INC) 28 June 2001 (2001-06-28) abstract page 6, line 9-20 figure 3	1,11 5
L	WO 03 049357 A (TELEFON AB LM ERICSSON) 12 June 2003 (2003-06-12) abstract page 6, line 6 -page 8, line 7 figures 1-3	1,10,11
E	EP 1 389 862 A (ALCATEL) 18 February 2004 (2004-02-18) abstract column 7-8, paragraphs 47-51 claims 1-6; figure 2	1-4,10, 11

BEST AVAILABLE COPY

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 03/50098

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 02093838	A	21-11-2002	WO	02093838	A1	21-11-2002
WO 0191374	A	29-11-2001	AU	6092601	A	03-12-2001
			EP	1299974	A1	09-04-2003
			WO	0191374	A1	29-11-2001
WO 02085041	A	24-10-2002	DE	10117998	A1	24-10-2002
			WO	02085041	A2	24-10-2002
			EP	1378108	A2	07-01-2004
WO 0147222	A	28-06-2001	AU	3265601	A	03-07-2001
			WO	0147222	A2	28-06-2001
WO 03049357	A	12-06-2003	GB	2376392	A	11-12-2002
			WO	03049357	A2	12-06-2003
EP 1389862	A	18-02-2004	EP	1389862	A1	18-02-2004

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 484 892 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
08.12.2004 Bulletin 2004/50

(51) Int Cl.⁷: H04L 29/06

(21) Application number: 04013351.4

(22) Date of filing: 07.06.2004

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PL PT RO SE SI SK TR
Designated Extension States:
AL HR LT LV MK

(30) Priority: 05.06.2003 US 455557

(71) Applicant: Nortel Networks Limited
St. Laurent, Quebec H4S 2A9 (CA)

(72) Inventor: Reith, Lothar
60596 Frankfurt (DE)

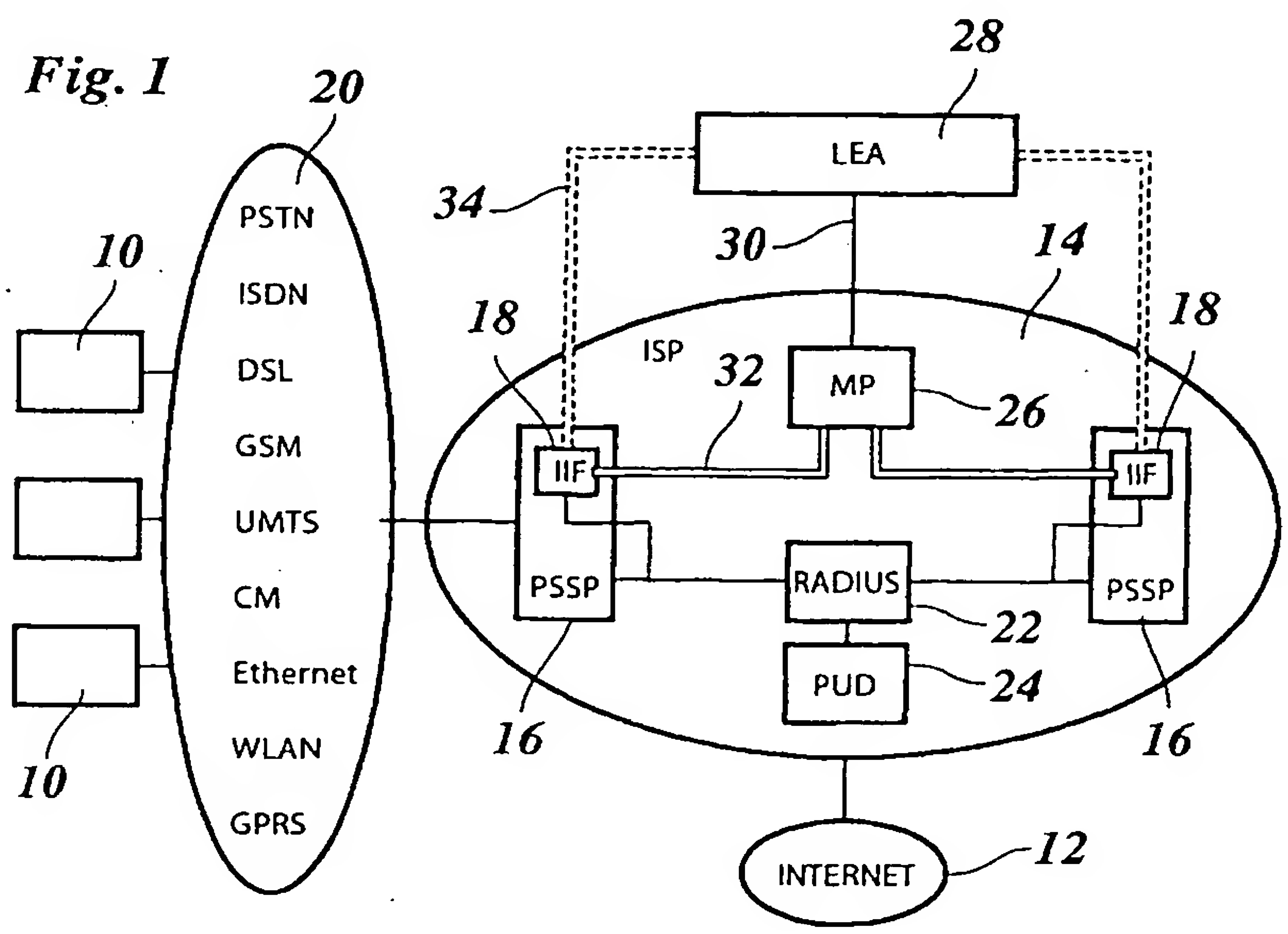
(74) Representative: Ameline, Jean-Paul et al
Nortel Networks S.A
Intellectual Property Law Group-CT34,
Parc d'activités de Magny-Chateaufort,
Chateaufort
F-78928 Yvelines Cedex 9 (FR)

(54) Method and system for lawful interception of packet switched network services

(57) A method for lawful interception of packet switched network services, comprising the steps of:

- when a user accesses the network and is identified by a target-ID at a primary interception point of the network, sending the target-ID to an interception management center,
- checking at the interception management center whether the user is a lawful interception target and sending an encrypted interception instruction set to a secondary interception point,
- decrypting said interception instruction set at the secondary interception point and performing an interception process in accordance with the interception instruction set, said interception process including the transmission of encrypted interception and dummy data to a mediation device, wherein said dummy data are added for obscuring true interception traffic between the secondary interception point and the mediation device.

Fig. 1



EP 1 484 892 A2

Printed by Jouve, 75001 PARIS (FR)

Description

BACKGROUND OF THE INVENTION

1. Field of the invention

[0001] The invention relates to a method and a system for lawful interception of packet switched network services.

[0002] According to recent legislation in many countries, providers of packet switched network services are obliged to provide facilities that permit lawful interception of the data traffic over the network. While some countries prescribe that all traffic of all users or subscribers to the network services shall be monitored, the laws of other countries provide that such general monitoring is forbidden and interception of traffic to or from users, even interception of only the connection data, is permitted only for specific users or subscribers who qualify, e. g. by court order, as lawful interception targets. Of course, the service provider has a responsibility to make sure that the identities of lawful interception targets are kept secret.

[0003] Accordingly, there is a demand for a method and a system for lawful interception of packet switched network services that can be implemented and operated at relatively low costs and can easily be adapted to differing legal provisions and requirements in various countries.

2. Description of the related art

[0004] A conventional approach is the so-called hardware monitoring, which means that specialized equipment necessary for interception purposes is installed at a location where the specified lawful interception target gets access to the network. This involves high costs and has the further drawback that the secrecy requirement is difficult to fulfill, because of the potential visibility of the hardware to not security-screened staff. Moreover, this approach is not practical when the network can be accessed from mobile units such as mobile telephones, laptop computers and the like, or through public access points such as WLAN hot spots or simply by dialing in over a PSTN with a modem or via ISDN from a hotel or public telephone.

[0005] Another known approach is the so-called software monitoring, wherein suitable software is implemented within the internal network of the service provider for identifying the subscribed users that connect to the network and for deciding whether or not the traffic to or from these subscribers shall be intercepted. This solution involves a certain amount of interception-related traffic within the internal network of the service provider, and this traffic may be observable by a relatively large number of employees of the service provider, so that careful security screening of the personnel is necessary in some countries. This not only constitutes a high cost factor but may also raise intricate legal problems in view of employment contracts and the like.

[0006] The European Telecommunications Standards Institute (ETSI) has published specifications for a lawful interception reference model (ETSI-document ES 201 671).

[0007] An Internet document of Baker et al.: "Cisco Support for Lawful Intercept in IP Networks", April 2003, http://www.rfc-editor.org/internet-drafts/draft-baker-slem-architecture-00.txt, recommends that intercept traffic between an interception point and a mediation device is encrypted in order to limit unauthorized personnel from knowing lawfully authorized intercepts.

SUMMARY OF THE INVENTION

[0008] According to the invention, a method for lawful interception of packet switched network services, comprises the steps of:

- when a user accesses the network and is identified by a target-ID at a primary interception point of the network, sending the target-ID to an interception management center,
- checking at the interception management center whether the user is a lawful interception target and sending an encrypted interception instruction set to a secondary interception point,
- decrypting said interception instruction set at the secondary interception point and performing an interception process in accordance with the interception instruction set, said interception process including the transmission of encrypted interception and dummy data to a mediation device, wherein said dummy data are added for obscuring true interception traffic between the secondary interception point and the mediation device.

[0009] A system implementing the method according to the invention comprises at least one Packet Switching Service Point (PSSP) that includes interception functionality (e.g. an Internal Intercept Function (IIF) as specified in the ETSI model) and thereby serves as the primary and/or secondary interception point, and a Mediation Device (MD) through which the intercepted data and related information are handed over to one or more Law Enforcement Agencies (LEAs) who want to receive and evaluate the intercepted data. The PSSP may be any node in the network where data packets, including packets that contain the user-ID of a subscriber to the network, can be intercepted. The above-mentioned primary and secondary interception points may be formed by different PSSPs but are preferably formed by one and the same PSSP. The system further comprises an Interception Management Center (IMC). This is the place where the interception policy is provisioned as request-

ed by the law enforcement agencies. The IMC stores the identities of lawful interception targets (user-IDs, device-IDs, access-line IDs or other means to identify a target user with reasonable probability), that are serviced by the one or more PSSPs that are associated to this IMC. The IMC may further store information on the modes and scopes of interception that are applicable to the various targets and non-targets.

[0010] As is well known in the art, a user who has subscribed to the services of a packet switched network service provider is uniquely identified by any suitable identification that is called "user-ID" and may consist of the name of the user or any other suitable identifier such as a pseudonym. Alternatively or additionally, a user or, more precisely, an interception target may be specified by an access line ID such as a telephone number, a DSL-Line-ID, an ATM virtual channel or the like. In the present application, the term "target-ID" is generic to user-IDs and access line IDs and device IDs such as the MAC-Adress of a network interface card utilized by the target user.

[0011] When a user starts a usage session he gets identified by a minimum of one target-ID. Sometimes multiple target-IDs are present. The following are common target-ID classes:

1. a User-ID (usually combined with password for authentication). This is often summarized as "something you know" (or at least are supposed to know - the user or a legitimate user may have stored the user-ID and the password on the device being used, so the current user may not need to know the user-ID if he has access to the device with username and password stored).

2. a Device-ID of a device that he is using (such as a MAC adress of a network interface card, or a mobile station ID of a mobile handset, or via a Subscriber Identification Module in a mobile phone). This class of target-ID may be summarized as: "something you own", and is particularly useful in mobile scenarios. An IP-adress such as an IP-Version 6 adress may be considered a device ID in a mobile IP scenario when the IP-adress is assigned to the device.

3. an access network resource ID referred to hereafter as access-line-ID. This is a network interface ID of a network element that is not owned by the user, rather by the service provider or a business-partner of the service provider. An example is a DSL-line ID in a DSL access network, or the combination of an ATM device name, slot-number, port-number and ATM virtual Circuit ID. Another example would be an IP-Adress permanently assigned to said network interface. This class of target-IDs may be summarized as: "something you probably utilize in the network" as is the case for example with the DSL-Line into the house of a target user. This concept is very similar to voice wiretapping in fixed networks, which is usually done to the telephone access line as well and intercepts all communications over that telephone line, regardless if the intended target user speaks or somebody else having access to the phone attached to the line.

[0012] When the user connects to the network with a target ID being a user-ID, a logon procedure is performed in which the user has to authenticate himself by indicating his user-ID and, optionally, a password and the like. Conventionally, this authentication process has the purpose to permit the service provider to check whether the user has actually subscribed to the services. In case of commercial service providers, the authentication process is also needed for billing purposes. In some cases the user identification or logon procedure is performed utilizing a device-ID for identification of the device used by the user, without requiring a password, for example when providing an IP address granting limited access via DHCP based on a MAC address presented by the device or by a network interface card being part of the device. Such procedure is common when providing limited scope access to a user prior to proper authentication. In case of fixed line access, there may be no special logon procedure, as the user is being considered fixed to a certain access line which may have been permanently provisioned with a fixed IP address for example (similar to the situation in telephony, where a telephone line is permanently provisioned with a fixed telephone number). In some cases the device may present an IP-address such as a fixed IP-Version 6 address that has been assigned to the device.

[0013] According to the invention, the fact that the user has to indicate his user-ID or utilize at least one target ID when connecting to the network is also utilized for interception purposes. To this end, the user ID (and the access line ID or device ID, as the case may be) is detected at the PSSP serving as an interception point. It will be clear that, in order to be able to intercept all subscribers to the network, if required, the PSSPs having interception facilities must be strategically located in the network so that no subscriber can get access without passing at least one interception point. The target-ID is sent to the IMC where it is checked against the list of lawful interception targets and explicit non-targets. The IMC responds to the same PSSP from which the target-ID originated - or else to another PSSP - with an encrypted message indicating at least whether or not the target-ID represents a lawful interception target. The response, which is called an interception instruction set, may further specify whether the target is identified by its user-ID (i. e. interception of traffic to or from this user) or by its access line ID (i. e. interception of all traffic over this line, irrespective of the identity of the user) or by another temporary target-ID that is included in the interception instruction set, and may also include additional

information. For example, the interception instruction set may include a "conditional interception instruction", instructing the PSSP to monitor the traffic associated with the target-ID and start the interception of the complete traffic or a portion of the traffic only when a certain trigger condition occurs, said trigger condition being one of: usage of certain network or content resources or usage of a certain catchword, virus signature or bit-pattern specified in the interception instruction set. As another example, the interception instruction set may specify different interception classes indicating whether all packets or only a random selection of packets or only a specified subset of packets originating from or sent to the target are to be intercepted. The PSSP will then intercept the data packets in accordance with these instructions and will send them, again in encrypted form, to the mediation device.

[0014] The PSSP includes both, encryption and decryption facilities. The IMC includes at least encryption facilities for the interception instruction set, and the mediation device includes at least decryption facilities.

[0015] It is an advantage of the invention that the traffic between the PSSP and the mediation device and also most of the traffic between the PSSP and IMC is encrypted, so that it cannot be understood by an observer monitoring the traffic (encryption of the target-ID sent to the IMC may however be dispensed with). Thus, even the service provider's employees, for whom it would most likely be possible to monitor the traffic, cannot easily discover the identity of the lawful interception target. From the viewpoint of secrecy requirements, it is a further advantage that it is not necessary to implement the functionality of the IMC at each individual PSSP. The IMC and the mediation device may be located remote from the PSSP(s) and may thus be centralized, so that considerable cost savings can be achieved without violating secrecy requirements. Further, since no information on the identity of the lawful interception targets is permanently present at the individual PSSPs, and, if present, is stored in encrypted form or in an encrypted file, the personnel having access only to the PSSPs will not be able to identify the interception targets or determine if a true interception target is accessing that particular PSSP. The identity of the interception targets will only be known to a very limited number of employees, if any, who have access to the information stored in the single IMC or relatively few centralized IMCs, or have special operator privileges not available to non-security screened staff. It is understood that only a few staff members of the service provider, if any, have access to a secured area or locked room where the IMC may be located as well as the Mediation Device.

[0016] According to another important feature of the invention the security and secrecy is further enhanced by obscuring even the fact that interception-related traffic occurs between the PSSP and the mediation device. To this end, the interception instruction set sent from the IMC to the PSSP may specify that even in those cases in which the user is not to be intercepted or is not even a lawful interception target at all, dummy data traffic is created between the PSSP and the mediation device, so that an unauthorized observer who may monitor the encrypted data traffic cannot decide whether the traffic he sees is only dummy traffic or a hint to an actual interception process.

[0017] This enables the service provider to outsource the operation of the IMC and/or the mediation device to a third party company, which may handle all interception warrants presented from law enforcement agencies on the service-provider's behalf, without any employee of the service provider knowing about the details of a warrant.

[0018] The dummy interception traffic may be triggered by real packet arrival events at the PSSP or, alternatively, by random events or any other events, such as timer expiry. However, the dummy traffic shall not contain any subscriber data. In case that real subscriber traffic was used as triggering event for the dummy traffic, the contents are scrambled and made useless, so that the receiver or an observer cannot gather any useful information on the actual subscriber traffic. Thus, in spite of the dummy traffic, the privacy of the subscriber will be protected in case that the subscriber is not a lawful interception target.

[0019] Optionally, the invention may further include one or more of the following features:

- Sending re-classification messages from the IMC to the PSSP in order to reclassify an already active user to a different interception mode when, for example, a new interception warrant has to be implemented for an already active user, a warrant for an active user shall be terminated when the duration of the warrant has expired, a warrant for an active user is being withdrawn prior to expiration, or when the scope of a warrant for an active user is being changed necessitating a reclassification, e.g. from partial to full interception, or from no-interception to dummy-interception, or from dummy interception to no-interception.

- Hiding the information about the user interception class associated with an active user from not security screened operations staff of the service provider, by implementing special operator command privileges at the PSSP, in order to prohibit non-intercept-privileged operators from being able to successfully execute commands that show the user interception class of an active user, and/or by storing the user interception class in encrypted form on the network elements, where the decryption key is not available to operators without intercept-privilege.

- Discarding the dummy data directly after receipt at the mediation device, or alternatively using these dummy data for obscuring handover traffic from the

mediation device to the law enforcement agency.

- Statically or dynamically determining at the IMC the relation between real interception traffic and dummy traffic considering both the cost of the dummy traffic as well as the security requirements under the circumstances, where the applied mix of user intercept classes may depend on the regulatory requirements mandated by authorities, the time of day, the amount of simultaneously active users at a specific interception point (PSSP), the current traffic load, the theoretical peak-bandwidth required for interception traffic of real targets from a specific interception point, risk classification levels associated with the operational model applied, and general risk levels prevailing over a period of time in a specific country as declared by governmental authorities.

[0020] In another embodiment of the invention a constant (or varying) amount of "camouflage" traffic is created and sent at all times (even if no real interception is taking place). This camouflage traffic is composed of true intercept traffic and dummy data at a ratio that depends on the demand for true intercept traffic, so that the true intercept traffic will always be hidden in the amount of camouflage traffic. The camouflage packets may have a fixed size or variable sizes that are unrelated to packet sizes used by a particular subscriber. The volume of the camouflage traffic will be at least as high as the maximum theoretical or practical volume of real interception traffic plus any overhead to encrypt and encapsulate it into the stream of fixed-length camouflage traffic packets. This would make it impossible for an observer performing traffic analysis to determine if a real interception is taking place, and it would make it totally impossible to determine the fact of lawful interception taking place, even when sending the internal lawful interception traffic to the MD over insecure public networks like the Internet. It would also make it impossible even for a malicious member of the operations staff (without interception operator command privileges) which is cooperating with a target, to test if a particular user is currently a target.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] Preferred embodiments of the invention will now be described in conjunction with the drawings, in which:

Fig. 1 is a diagram illustrating a system according to one embodiment of the invention;

Fig. 2 and 3are diagrams illustrating two examples of the method according to the invention;

Figs. 4 to 9 are diagrams showing a modified embodiments of the system; and

Fig. 10 illustrates a method of combining intercept traffic with dummy traffic.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0022] As is shown in figure 1, a packet switched network services provider, an Internet Service Provider (ISP) in this example, has responsibility for a certain number of facilities allowing a number of end users 10 to get access to the network, i.e. the Internet 12. These facilities are interconnected by an internal network 14 of the ISP and comprise a number of Packet Switching Service Points (PSSP) 16, i.e. switching nodes, that are each equipped with an Internal Interception Function (IIF) 18.

[0023] In the example shown, the PSSPs 16 equipped with the IIFs 18 are situated at the subscriber edge of the network 14, i.e. the place where the end users 10 connect to the internal network 14 and hence to the Internet 12 via any suitable access network 20 such as a Public Switched Telephone Network (PSTN), an integrated Services Digital Network (ISDN), a Digital Subscriber Line (DSL) access network, a mobile telephone network (2G like GSM, 2,5G like GPRS or 3G like UMTS), a WLAN access network, an Ethernet access network or a Cable Modem access network (CM) or a combination of the same. However, the PSSPs may also be located at any other node within the internal network 14, as long as it is assured that the target data traffic of interest to and from the end users 10 will pass at least one of the PSSPs equipped with an IIF 18. As an example, a PSSP may be a "Shasta 5000 BSN" (trademark) available from Nortel Networks Limited (BSN stands for Broadband Services Node). Through the internal network 14, the PSSPs are connected to at least one authentication server, in this example a_"Remote Authentication Dial-In User Service" (RADIUS) server 22, co-operating with a Personal User Data Base (PUD) 24 which stores the user data of the subscribers (the RADIUS protocol is described in RFC 2865, entitled "Remote Authentication Dial-In User Service (RADIUS)", and in RFC 2866 entitled "RADIUS Accounting", both published by the Internet Engineering Task Force organization (IETF) in June 2000).

[0024] When an end user 10 connects to the services of the ISP, he will authenticate himself by a suitable user-ID by which the specific user is uniquely identified. The PSSP 16 forwards the user-ID to the RADIUS server 22, thereby triggering an authentication procedure in which the user-ID is checked against the personal user data base 24 to see whether the user is authorized to the services of the ISP. When the authentication procedure is successful, a user session for this specific user starts, and the user may be recorded in the personal user data base 24 as an active user. When the user logs off or gets disconnected from the PSSP, the user may again be stored as an inactive user. The messages in-

dicating the start and the end of a user session will be stored and processed for billing purposes if the user has not subscribed to a flat rate.

[0025] The internal network 14 further comprises at least one Mediation Point (MP) 26 which serves as an interface between the internal network 14 of the ISP and a Law Enforcement Agency (LEA) 28 that is authorized to intercept the traffic of either all users or of a number of specified users that qualify as lawful interception targets. The identities of the lawful interception targets are stored at the mediation point 26, preferably together with more detailed information on the mode and scope of interception that is allowed and desired for each individual target. The mediation point 26 is connected to the facilities of the law enforcement agency 28 through a safe communication channel 30 which may be used for sending the intercepted data to the LEA 28 and also for loading the information specifying the interception targets into the mediation point 26.

[0026] Through the internal network 14, the mediation point 26 is connected to the interception function 18 of at least one, preferably a plurality of PSSPs 16, as is symbolized by broad, contoured connection links 32 in figure 1. The contoured representation of the links 32 indicates that traffic on these links occurs only in encrypted form.

[0027] When an end user 10 has logged on by the procedure described above, the user-ID that is sent to the RADIUS server 22 is also supplied to the internal interception function 18 of the pertinent PSSP 16. Triggered by this event, the IIF 18 creates an encrypted interception instruction request, including the encrypted user-ID, and sends the same via link 32 to the mediation point 26. Here, it is checked whether the user who has logged on is a lawful interception target, and an encrypted response is sent back to the IIF 18 through the link 32. This encrypted response message indicates whether or not the user is to be intercepted and in which way this is to be done. In accordance with the instructions contained in this encrypted response, the IIF 18 will intercept some or all of the traffic from or to the end user 10 and will send the intercepted data and/or intercept related information, again in encrypted form, to the mediation point 26 from where they are forwarded to the law enforcement agency 28 through the safe channel 30. As an alternative, the intercepted and encrypted data may be sent directly to the law enforcement agency 28 through encrypted channels 34, as has been indicated in phantom lines in figure 1.

[0028] An example of such an interception procedure will now be described by reference to figure 2. In step S1, a user 10 logs on to the services provided by the ISP and is identified by a target-ID, a user-ID in the present example. In step S2, the PSSP 16 through which the user has connected to the network, or more precisely the IIF 18 thereof, sends the encrypted user-ID to the mediation point 26. In step S3, the mediation point 26 returns an encrypted lawful interception instruction set to the PSSP 16. This instruction set includes at least the information that the user shall be intercepted or shall not be intercepted. Instructions may further specify other intercept related information, for example, that only access-connection data (e.g. time and duration of the user's online-usage session) or only certain end to end connection data (e.g. URLs of websites visited, or IP addresses of Voice over IP communication partners) but not the contents of the communications itself shall be intercepted. Another instruction may specify that all traffic (connection data and/or contents) to and from the user shall be intercepted or only messages sent from the user to another destination or only messages sent from other sources and received by the user. Yet another instruction may specify that all data packets or only a subset of the transmitted data packets (e.g. a random selection) shall be intercepted or that interception of all following data packets shall be triggered by specific data packets that represent specific catch words that are related to unlawful activities. Yet another instruction may specify that interception is restricted to traffic to or from specific sites or classes of sites, e.g. web servers located in a specific country, or to specific protocols or flows such as SIP traffic and RTP traffic which are utilized to signal and carry voice over IP or multimedia communications.

[0029] The internal interception function 18 will then perform the interception procedure in accordance with these instructions. In step S4, the user connects to a web site in the Internet 12, typically by entering a Universal Resource Locator (URL) of the desired web site. Then, in step S5, the connection data, i.e. the URL, will be sent in encrypted form to the mediation point 26.

[0030] If the instruction set specifies that contents shall also be intercepted, the data packages representing the contents of the selected web page and being sent to the user 10 will also be intercepted and will be sent in encrypted form to the mediation point 26 or to the LEA 28 in step S6.

[0031] As another example, the steps S4-S6 may also consist of the user 10 sending an e-mail to a specific e-mail address. Then, the encrypted e-mail address will be transmitted in step S5 and the encrypted contents of the e-mail will be transmitted in step S6. Conversely, if step S4 consists of the user retrieving an e-mail from his mail box, steps S5 and S6 will consist of encrypting and transmitting the origin and the contents of the e-mail. If the mail box of the pertinent user is provided by a foreign ISP in another country, this mail box may also be guarded by a PSSP having an internal interception function 18 and located at a border gateway, so that the e-mail addressed to the specific user may be intercepted already when it is sent to the mail box.

[0032] In step S7, the user logs off or disconnects from the internal network 14 of the ISP. This triggers an encrypted log off message being sent to the mediation point 26 in step S8.

[0033] It will be understood that, because all the traffic

between the PSSP 16 and the mediation point 26 is encrypted, this traffic can only be understood by the pertinent equipment and not by any individuals monitoring the traffic on the channel 32, not even by the personnel of the ISP itself, except the very restricted number of employees having access to the mediation point 26. Thus, secrecy of the interception-related information can be assured with high reliability. Since all relevant interception-related instructions are stored centrally in the mediation point, the system can easily be managed at low costs. The hardware and software components of the internal interception functions 18 to be implemented in the individual PSSPs 16 are the same for all PSSPs.

[0034] Figure 3 illustrates the method that is employed in cases where the user who has logged on in step S1 is not to be intercepted at all. In this case, the response to the request S2 in step S3' consists of a dummy traffic command specifying that the user is not to be intercepted but dummy traffic shall be generated on the encrypted link 32 in order to disguise the fact that this user is not being intercepted. This will make it difficult for a person monitoring the traffic on the link 32 to draw any conclusions as to the identity of lawful interception targets from the traffic occurring on this link.

[0035] The dummy traffic may be generated by the interception function of the PSSP 16 at random. In the embodiment shown in figure 3, however, this traffic is also triggered by the events S4 and S7 and by the occurrence of data packets to or from the user at the PSSP 16. Thus, when the user has connected to a web site in step S4, this event triggers encrypted dummy traffic in step S5'. The contents of this traffic will however be senseless or scrambled and in any case anonymized, so that the law enforcement agency or an observer cannot gain any knowledge on the actual event S4. It may be discarded at the mediation point directly upon receipt. Thus, this kind of traffic will be allowed even in cases where interception of the pertinent user is legally forbidden. Similarly, any packet events at the PSSP 16 will trigger encrypted dummy traffic in step S6' in order to mock the interception of contents. Of course, such dummy traffic may also be generated in case of figure 3 if the lawful interception instruction set specifies intercept related information, e.g. that only connection data but no contents are to be intercepted. Further, the dummy traffic command sent in step S3' may itself include senseless "dummy" data in order to make the length of this command resemble the length of a true interception instruction set.

[0036] When, in figure 3, the user has logged off in step S7, this triggers an encrypted dummy termination command in step S8' mocking the step S8 in figure 2. Since, however, the identity of the user is not known to the LEA 28 or to an observer, no meaningful information can be gathered from the step S8', neither.

[0037] Although the system is capable of real time interception, it may be advantageous to send the messages in steps S5, S5' and S8, S8' with a random time delay, so that the user may not be identified through coincidence of events S4 and S5 or S7 and S8. The exact time of the events S4 and S7 may be included in the encrypted messages in the form of a time stamp, if the user is a lawful target.

[0038] Comparing figures 2 and 3, it can be seen that, unless the encryption code is cracked, the pattern of traffic on the link 32 for users that are actually being intercepted is indistinguishable from the pattern for users that are not intercepted.

[0039] Since all the traffic on the link 32 is encrypted, the mediation point 26 may even be located outside of the internal network 14 of the service provider. This has been exemplified in figure 4, where the mediation point 26 is located within the facilities of the law enforcement agency 28. In some countries, it may however be required that the service provider has control over the mediation point 26. In other countries, it may be required that the mediation point is located in the domain of the Law Enforcement Agency, in yet other countries it may be mandated or at least possible that the mediation point is being operated by a third party that is especially certified by governmental authorities.

[0040] The mediation point 26 may store the target-IDs of all active users together with an identification of a minimum of one PSSP used for accessing the network, and an identifier used to identify the usage session within that PSSP, so that the interception of a new target may be provisioned by sending an appropriate interception instruction set even when the user is already active. Likewise, the interception may be terminated or the interception instruction set may be changed while the user remains active.

[0041] Figure 4 further shows an example of a PSSP 16' for which the interception function (IF) 18 is not internal to the PSSP but is implemented in a device outside of the PSSP and connected thereto by a suitable interface.

[0042] As is shown in figure 5, the function of the mediation point 26 can be subdivided into two main function blocks which are called Intercept Management Center (IMC) 36 and Mediation Device (MD) 38. The IMC 36 is the function that receives the user ID or, more generally, the target-ID form the IIF 18 and returns the interception instruction set IIS. The MD 38 is the entity that receives the encrypted intercept data and/or dummy data from the IIF 18 and implements the handover interface to a Monitoring Center (MC) 40 in the law enforcement agency 28. If the line 30 connecting the MD 38 to the MC 28 is not considered to be safe enough, the data handed over to the Monitoring center 40 may still include the dummy data generated by the IIF 18.

[0043] Figure 6 shows a modified embodiment, in which the interception management center 36 and the mediation device 38 are not integrated into a common device (such as the mediation point 26 in figure 5) but are embodied as separate physical entities. In this case

the PSSP 16, the IMC 36, the MD 38 and the MC 40 might be operated by two, three or even four different legal entities.

[0044] According to a modification which has not been shown, the mediation device (MD) 38 might as well be combined with the monitoring center (MC) 40 in the LEA 28.

[0045] Figures 7 to 9 show different arrangements of the interception management center (IMC) 36 in relation to the RADIUS Server 22 and the PSSP 16. In figure 7 the IMC 36 acts as a "proxy RADIUS server". This means that the IMC appears as a RADIUS server toward the PSSP 16 which acts as a RADIUS client, and at the same time the IMC acts as a RADIUS client towards the RADIUS server 22. The traffic between these three entities is governed by the RADIUS protocol.

[0046] In figure 8, so function of the IMC has been incorporated in the RADIUS server 22. In figure 9, the line interconnecting the RADIUS server 22 and the PSSP 16 includes a tapping device 42 which is capable of intercepting and manipulating RADIUS messages. RADIUS response messages from the RADIUS server 22 towards the PSSP 16 are manipulated by the tapping device 42 either by manipulating an interception instruction set that is already present in the RADIUS message or by inserting a new interception instruction set under the control of the IMC 36. Tapping device 42 may for example be formed by a web switch "ALTEON" (trademark) supplied by Nortel Networks Limited.

[0047] Figure 10 illustrates another embodiment of the method for obscuring the traffic between the IIF 18 and the mediation device (MD) 38 and possibly also between the MD 38 and the MC 40. Here, the traffic consists of a continuous stream of encrypted "camouflage" packets 44 of a fixed size that are constantly transmitted from the interception point (PSSP) to the mediation device, regardless of whether or not or how much true interception traffic is generated by PSSP. If there is no interception traffic at all, the camouflage packets 44 consist only of dummy data. Conversely, if the volume of true interception traffic reaches the capacity limits of the continuous stream of the camouflage packets 44, these packets are almost completely filled up with intercepted data.

[0048] The top line in figure 10 illustrates an intercepted data packet that has to be transmitted to the mediation device 38 and, in the example shown, has a length greater than the transport capacity of a single camouflage packet 44. Then, the contents of the intercepted packet 48 are distributed over a sufficient number of camouflage packets 44 (two in the given example), as is shown in the second line in figure 10. This line shows the format of transport packets, 50, 52 and 54 that are to be converted into the camouflage packets 44 through encryption. Each transport packet includes a minimum of one fragment-header, which contains at least a significance bit 56. If this bit is set to "0", then the remainder of the transport packet contains only dummy traffic (64, 66). If this bit is set to "1", the fragment header also contains, an interception ID 57, which identifides the current user-session of the target, a length field 58 and a "more" bit 60. The header - if significant - is followed by a fragment load section 62, which in case of the fragment load 62 that is contained in transport packet 50 is identical to the maximum load section of the transport packet and thus to the maximum transport capacity of a single camouflage packet. In case of the transport packet 50, the fragment load section 62 is filled to its full capacity with a first fraction 48a of the intercepted packet 48. The significance bit 56 indicates that the contents of the fragment load section 62 are significant, i. e. represent true intercepted data. The "more" bit 60 indicates that fragmentation has occurred and that the subsequent fragment load section 62 includes only a fragment of the intercepted packet 48 which will be continued in the next transport packet 52. If the intercepted packets and/or an initial fragment of a packet 48 are relatively short, it is possible that two or more intercepted packets are included in multiple fragment load sections 62 contained in a single transport packet. Then each data packet_or fragment has its own fragment header, as a single fragment load section 62_can also carry a full packet if it is sufficiently short. The length field 58 of the fragment header indicates the length of the corresponding fragment load section 62.

[0049] In the transport packet 50, the significance bit 56 is "1", because the fragment_load section 62 carries the first fragment of the intercepted packet 48, and the "more" bit 60 is also "1", because another fragment 48b of the packet 48 will be included in the next transport packet 52.

[0050] In case of the transport packet 52, the significance bit 56 is "1", but a "more" bit 63 is "0", because this transport packet will include all the rest of the current intercepted packet 48. The fragment load section 62 of packet 52 includes the last fragment 48b of the intercepted packet 48, and the length of this fragment is indicated in a length field 61. Each fragment-load section is immediately followed by a next fragment header, if the fragment has not filled the transport capacity completely. In case of packet 52, another fragment header follows which consists only of the significance bit 56 (set to "0"), which means that the remainder of the transport packet is insignificant and carries only meaningless dummy data 64. However, multiple fragment sections 62 could have followed instead of dummy data 64, carrying short full packets and the last fragment section could have carried an initial fragment of a larger packet not fully fitting within the remainder of the transport packet 52.

[0051] Since, in the present example, no further intercepted packet needs to be transmitted, the next transport packet 54 has a header consisting only of the significance bit 56 with the value "0" which is consequently followed by an insignificant fragment section 66 in this case.

[0052] After the transport packets 50, 52, 54 have

been encrypted to form the camouflage packets 44, it is impossible for an observer doing traffic analysis to decide whether or not true interception traffic occurs.

[0053] The length and/or the transmission frequency of the camouflage packets 44 may be varied in accordance with the overall traffic load on the network, in order to make sure that there will always be a sufficient transport capacity for the true interception traffic.

[0054] In a modified embodiment, in order to allow for variable length camouflage packets 44, the first significance bit in a camouflage packet may be replaced by a significance field, which comprises the significance bit followed by the total length_of the transport packet (also implicitly defining the length of the camouflage packet 44, as depending on the encryption algorithm used, the lengths of the transport packet and of the camouflage packet would normally be the same).

Claims

1. A method for lawful interception of packet switched network services, comprising the steps of:

   - when a user accesses the network and is identified by a target-ID at a primary interception point of the network, sending the target-ID to an interception management center,
   - checking at the interception management center whether the user is a lawful interception target and sending an encrypted interception instruction set to a secondary interception point,
   - decrypting said interception instruction set at the secondary interception point and performing an interception process in accordance with the interception instruction set, said interception process including the transmission of encrypted interception and dummy data to a mediation device,

   wherein said dummy data are added for obscuring true interception traffic between the secondary interception point and the mediation device.

2. The method of claim 1, wherein said secondary interception point is identical to said primary interception point.

3. The method of claim 1, wherein the dummy data are generated at random.

4. The method of claim 1, wherein the dummy data are based on actual traffic to or from the pertinent user, but this traffic is scrambled such that, even after decryption, the contents thereof may not be reconstructed at the mediation device.

5. The method of claim 1, comprising a step of sending a continuous stream of camouflage packets from the secondary interception point to the mediation device, said camouflage packets including intercepted data in accordance with the demand and being filled up with dummy data to their full length.

6. The method of claim 1, wherein the interception instruction set includes a "conditional interception instruction", instructing the PSSP to send intercept related information or to monitor the traffic associated with the target-ID and start the interception of the complete traffic or a portion of the traffic only when a certain trigger condition occurs, said trigger condition being one of:

   usage of certain network or content resources or usage of a certain catchword, virus signature or bit-pattern specified in the interception instruction set.

7. A system for carrying out the method as claimed in claim 1, comprising:

   at least one interception point formed by a node in the network,

   an interception management center, and

   a mediation device serving as an interface between the network and a law enforcement agency for which interception services are provisioned,

   wherein said at least one interception point is adapted to send a target-ID of a user accessing the network to said interception management center,

   the interception management center is adapted to send to at least one of said interception points an encrypted interception instruction set to be decrypted at the interception point and enabling the same to perform an interception process in the course of which intercepted data are encrypted and sent to said mediation device, and

   the at least one interception point is further adapted to generate dummy data and to encrypt and send either the intercepted data or the dummy data or a combination of these, such that the occurrence of intercepted data is obscured.

8. The system of claim 7, wherein the at least one interception point is formed by a node of the network that is situated at a subscriber edge of the network, where end users connect to the network.

9. The system of claim 7, wherein the interception point is a switch adapted to connect end users to an IP or Ethernet network.

10. The system of claim 8, wherein the interception point is a switch adapted to connect end users to an IP or Ethernet network.

11. The system of claim 7, comprising a plurality of interception points connected to the same interception management center.

12. The system of claim 7, wherein said interception management center contains means for communicating with said PSSP according to the RADIUS protocol, and means for intercepting RADIUS messages either directly or using a tapping device (42) in a way that is transparent to a RADIUS server.

13. The system of claim 7, wherein said interception management center contains means for communicating with said PSSP according to the RADIUS protocol, and means for acting as RADIUS proxy server towards the client PSSP and a RADIUS server.

14. The system of claim 7, wherein said interception management center is combined with a RADIUS server.

Fig. 1

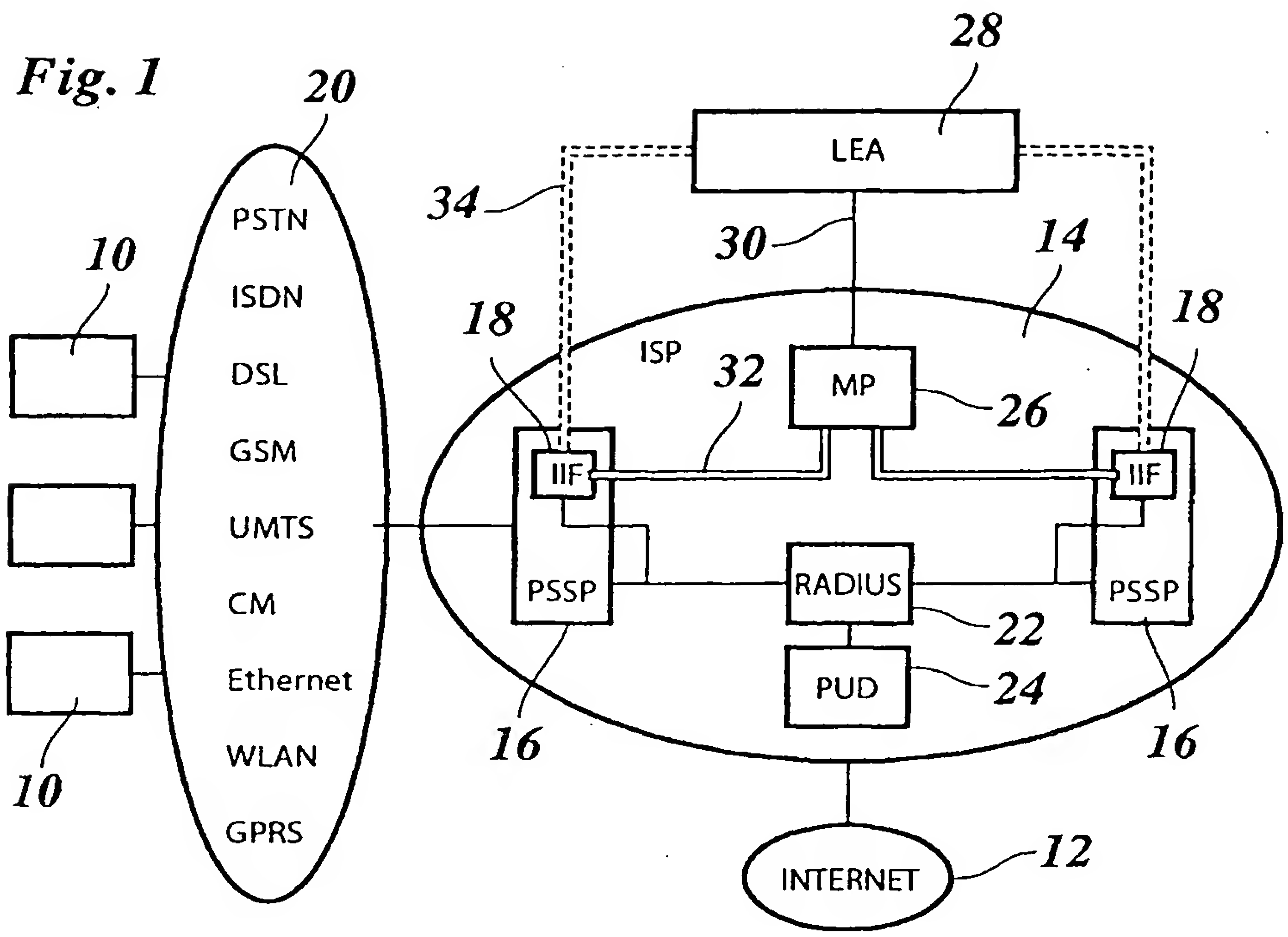

28
LEA
34
30
14
20
PSTN
ISDN
DSL
GSM
UMTS
CM
Ethernet
WLAN
GPRS
10
10
18
ISP
32
MP
26
18
IIF
IIF
PSSP
RADIUS
PSSP
22
PUD
24
16
16
INTERNET
12

Fig. 2

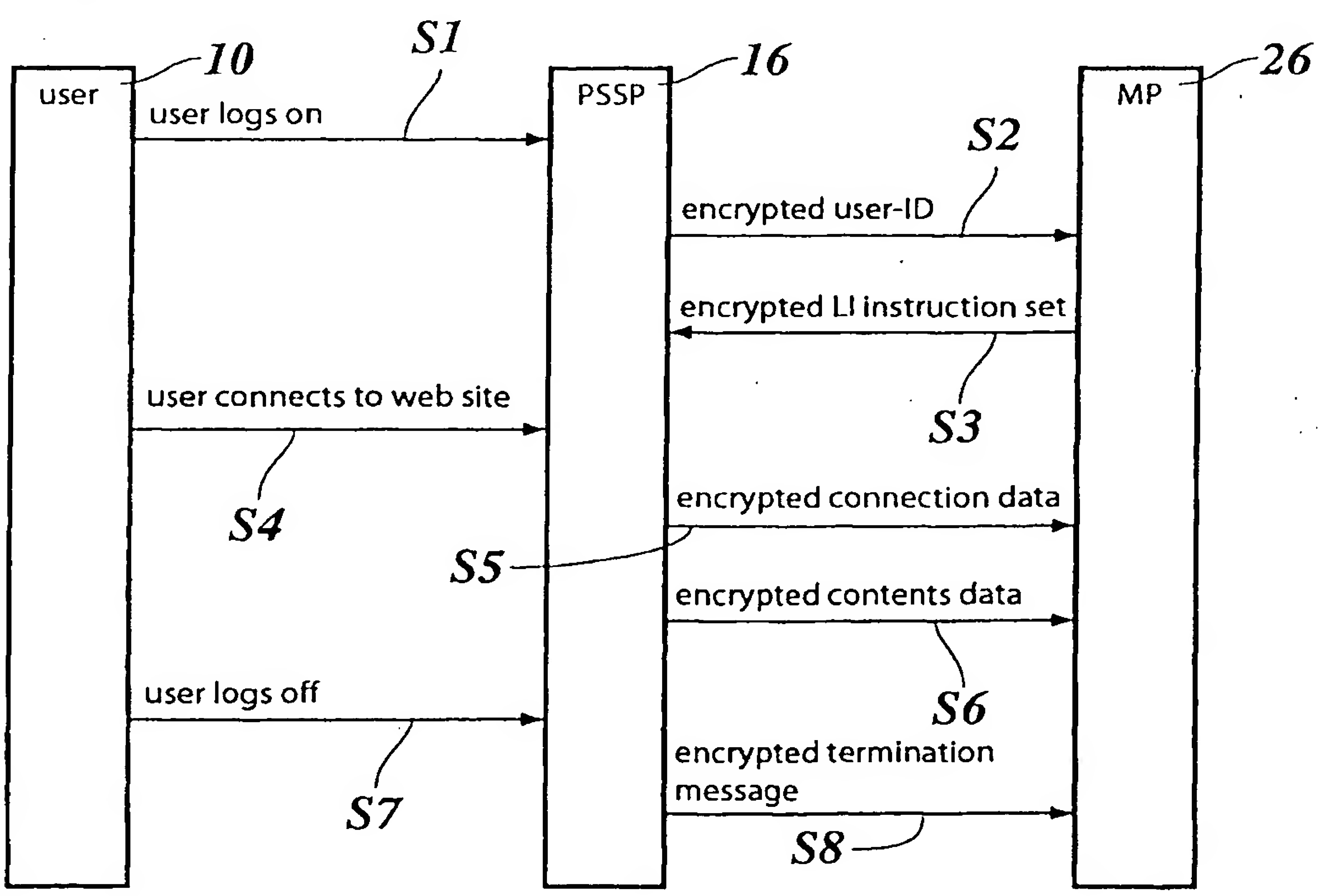

user
10
S1
user logs on
PSSP
16
MP
26
S2
encrypted user-ID
encrypted LI instruction set
S3
user connects to web site
S4
encrypted connection data
S5
encrypted contents data
S6
user logs off
encrypted termination message
S7
S8

Fig. 3

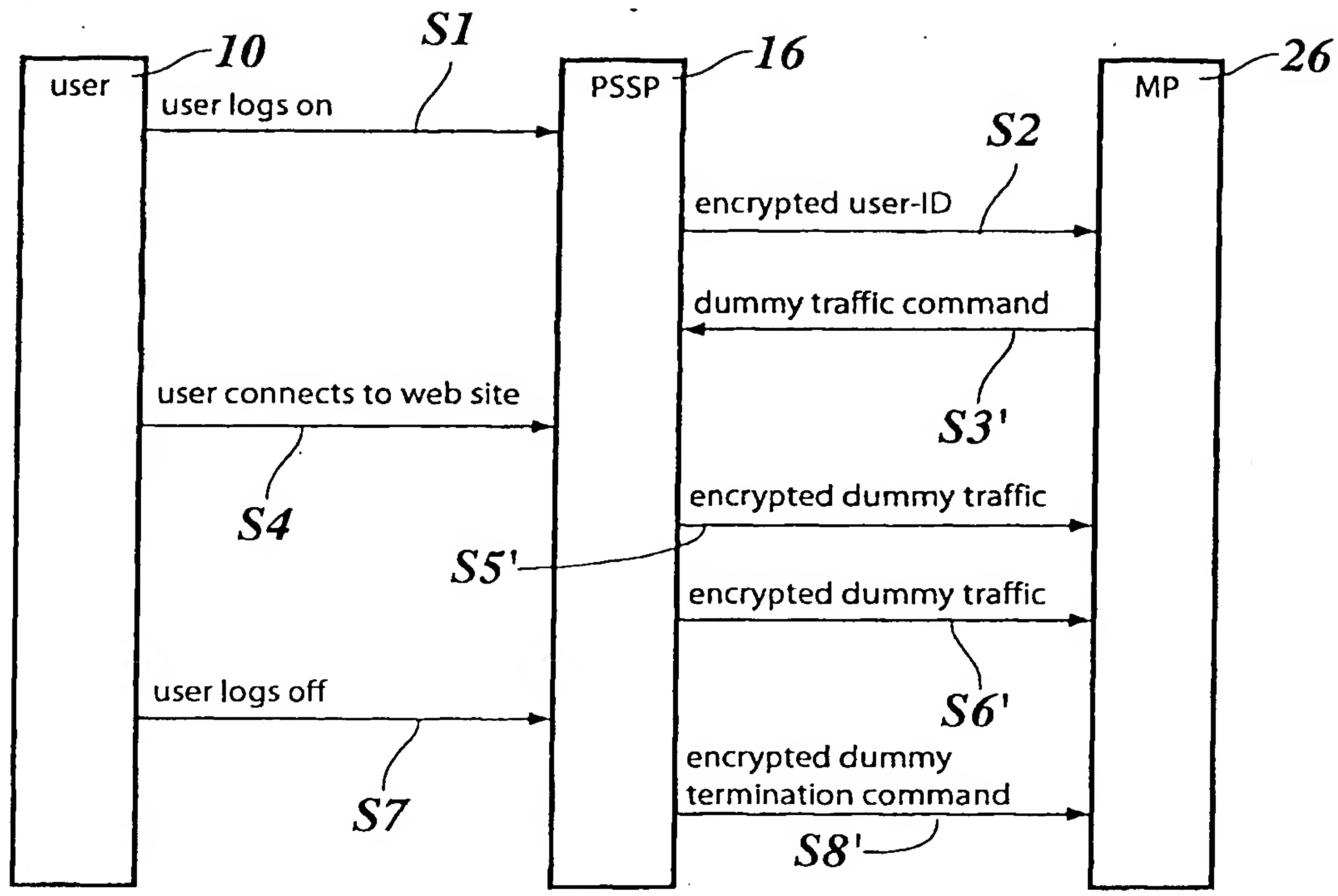

user
10
S1
user logs on
PSSP
16
MP
26
S2
encrypted user-ID
dummy traffic command
S3'
user connects to web site
S4
encrypted dummy traffic
S5'
encrypted dummy traffic
S6'
user logs off
S7
encrypted dummy
termination command
S8'

Fig. 4

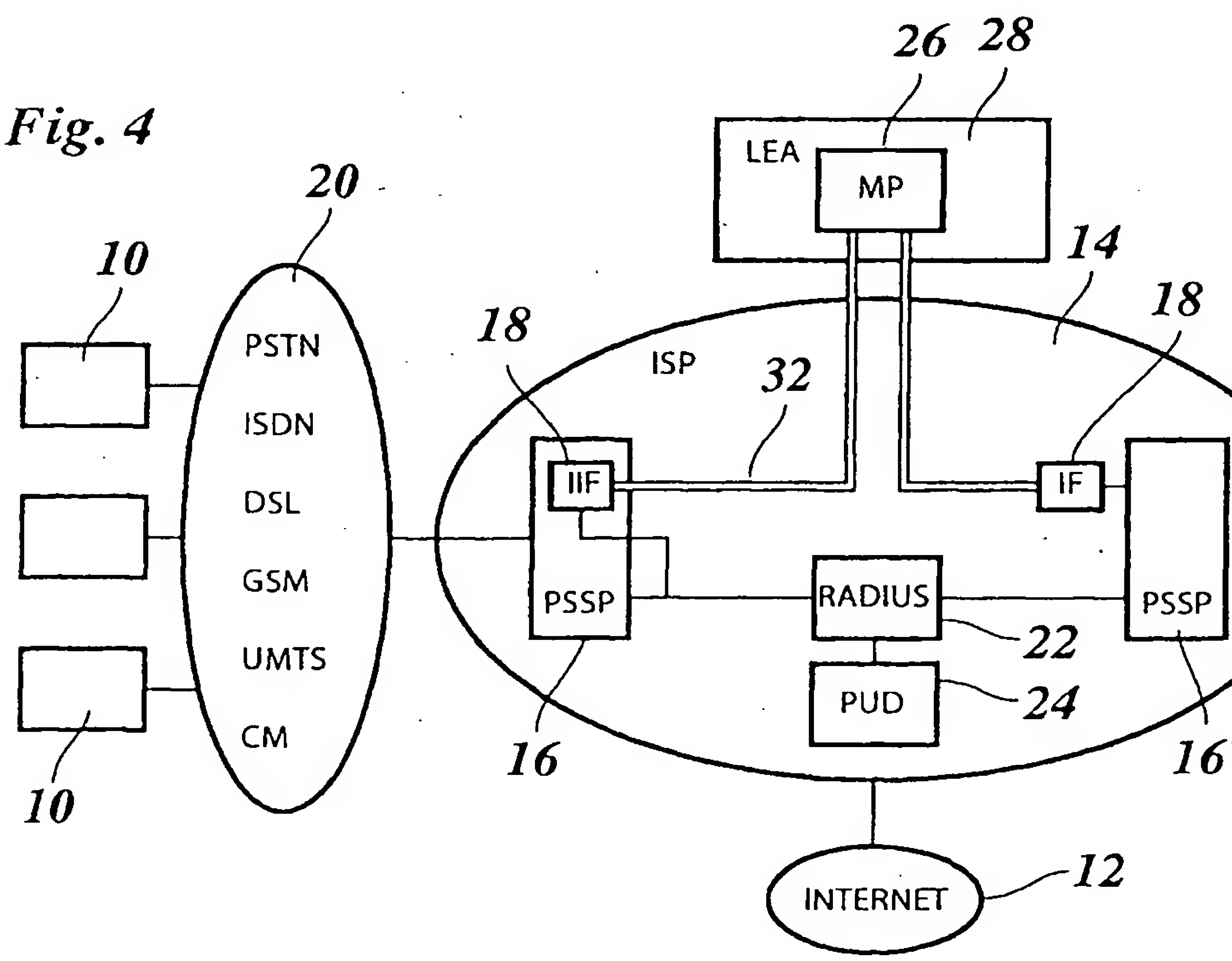

26
28
LEA
MP
20
10
14
18
PSTN
ISDN
DSL
GSM
UMTS
CM
10
18
ISP
32
IIF
IF
PSSP
RADIUS
PSSP
22
PUD
24
16
16'
INTERNET
12

Fig. 5

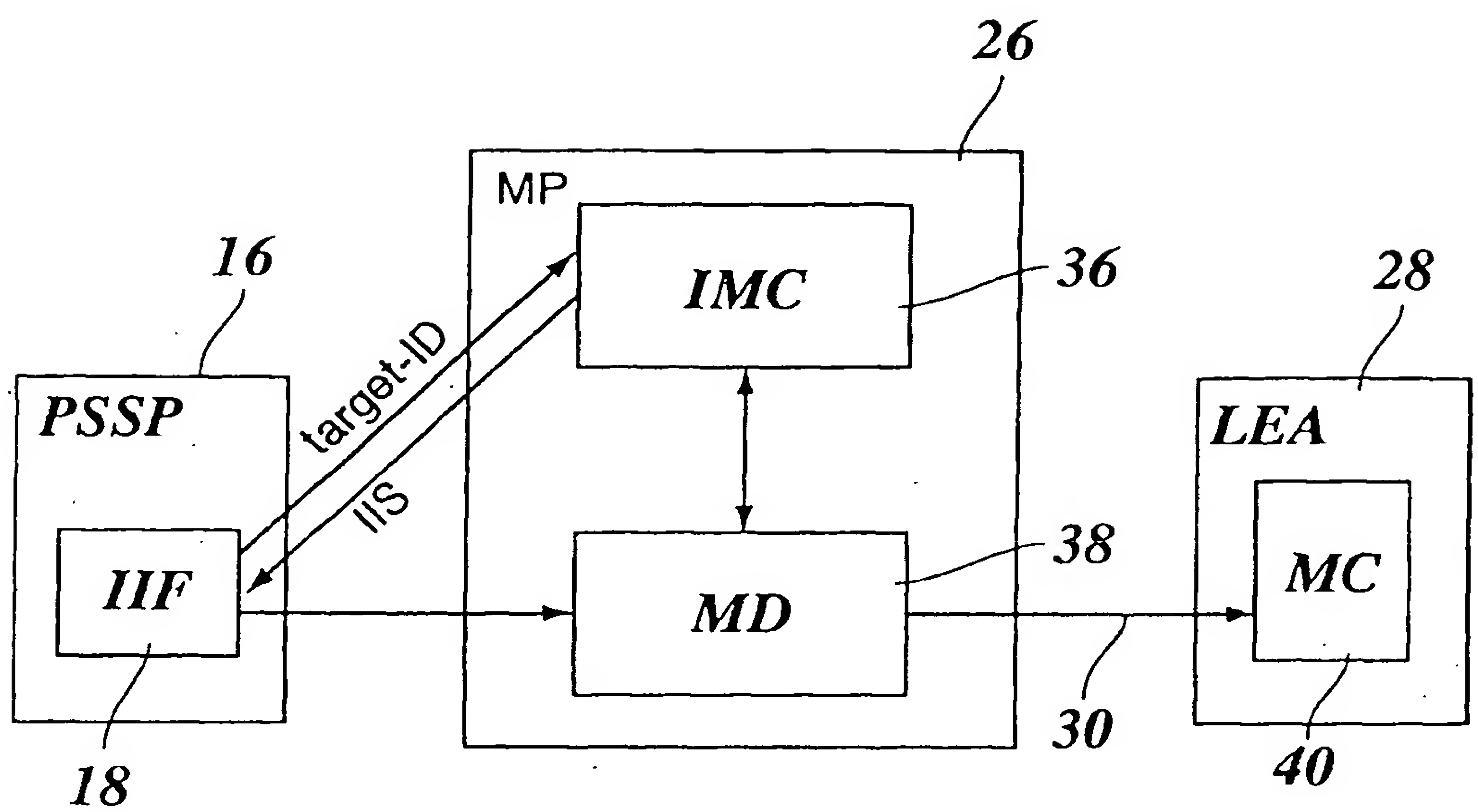

26
MP
IMC
36
16
28
target-ID
PSSP
LEA
IIS
38
IIF
MD
MC
30
18
40

Fig. 6

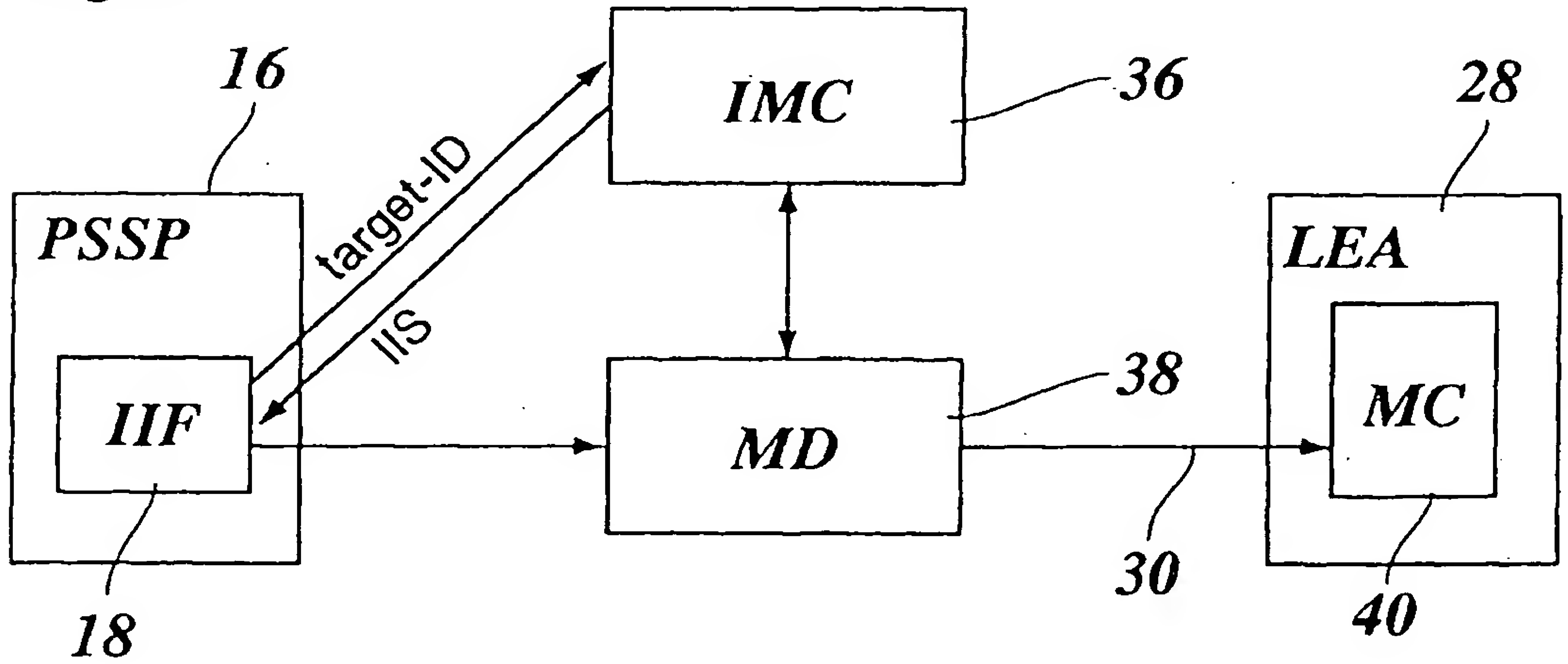

16
IMC
36
28
target-ID
PSSP
LEA
IIS
38
IIF
MD
MC
30
18
40

Fig. 7

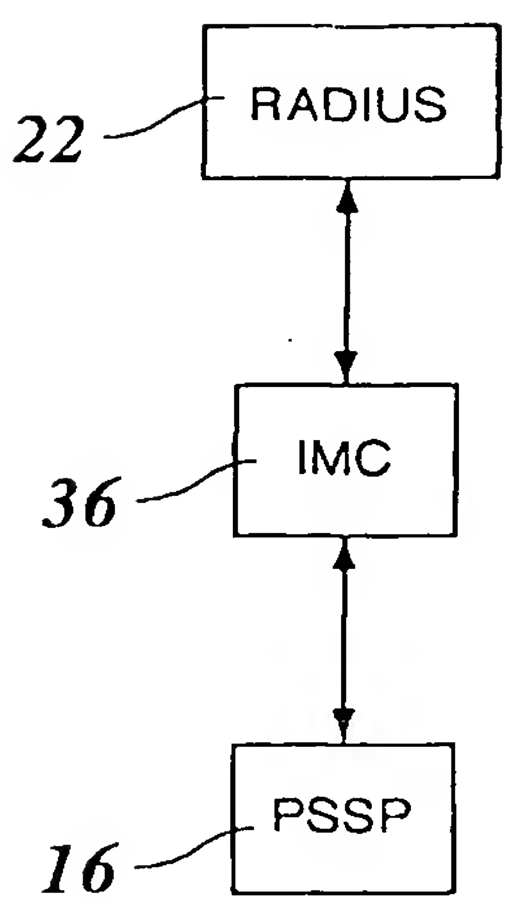


Fig. 8

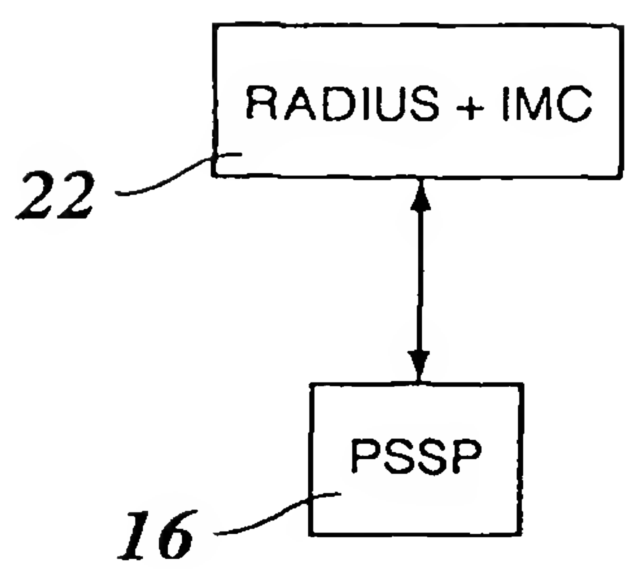


Fig. 9

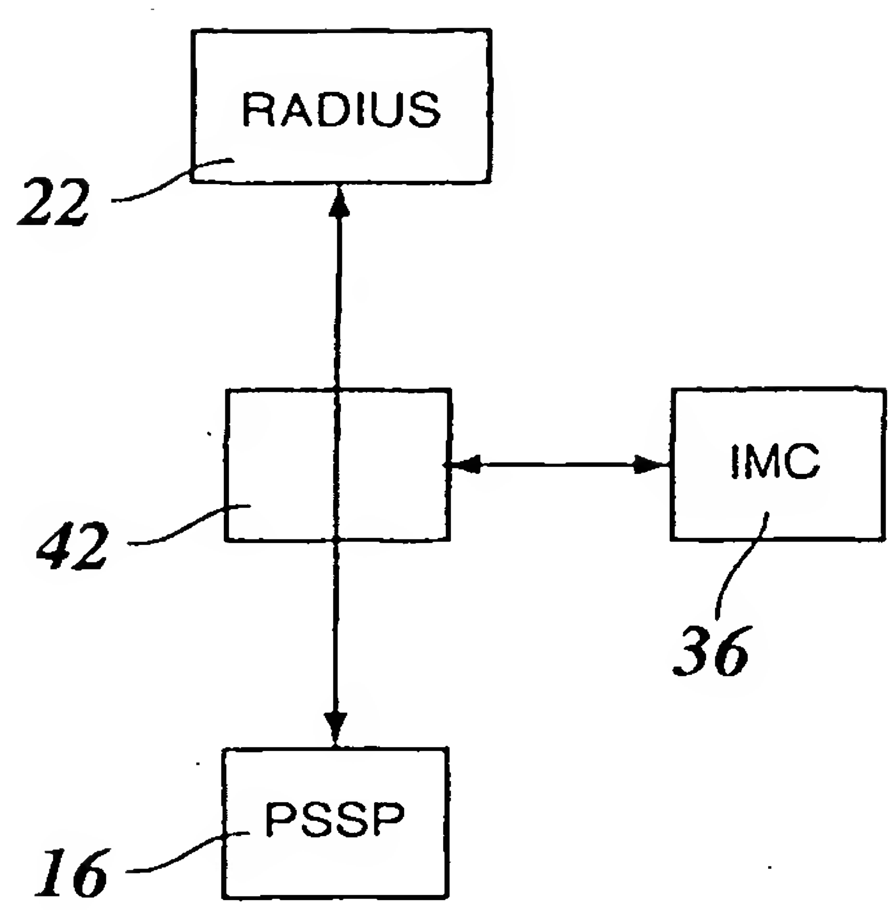
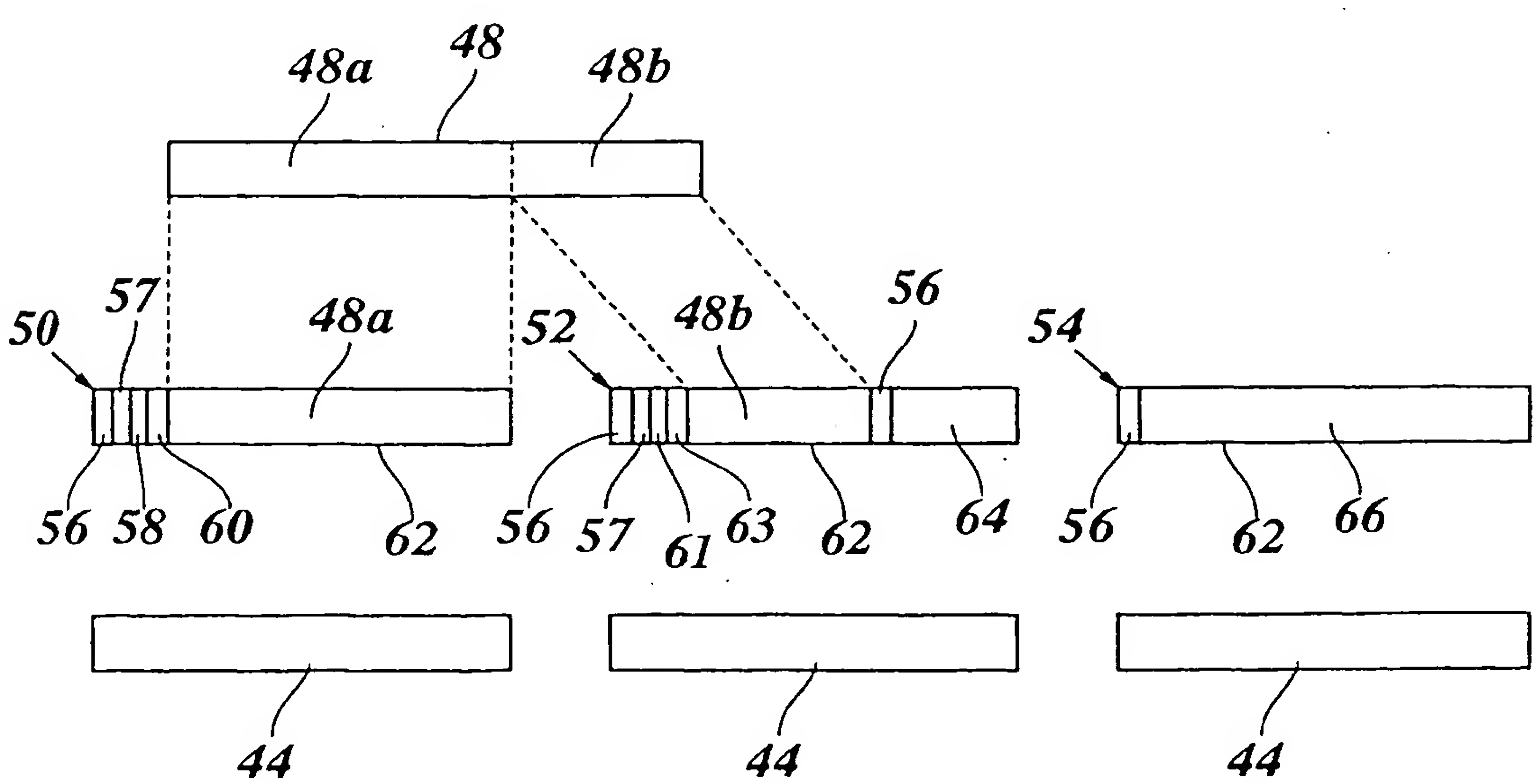


Fig. 10



(19)
Federal Republic of Germany
German Office of Patents and Trademarks

(10) **DE 103 58 333 A1** 2005.07.14

(12)

Document of Disclosure

(21) Reference: **103 58.333.5**
(22) Date of registration: **Dec.12, 2003**
(43) Date of disclosure: **July 14, 2005**

(51) Int Cl.[7]: **H04M 3/22**
G10L 17/00

(71) Registering company
Siemens AG, 80333 Munich, Germany

(72) Inventors:
Schmöller, Franz, 82110, Germering, Germany
Zimmermann, Beate, 82065, Baierbrunn, Germany

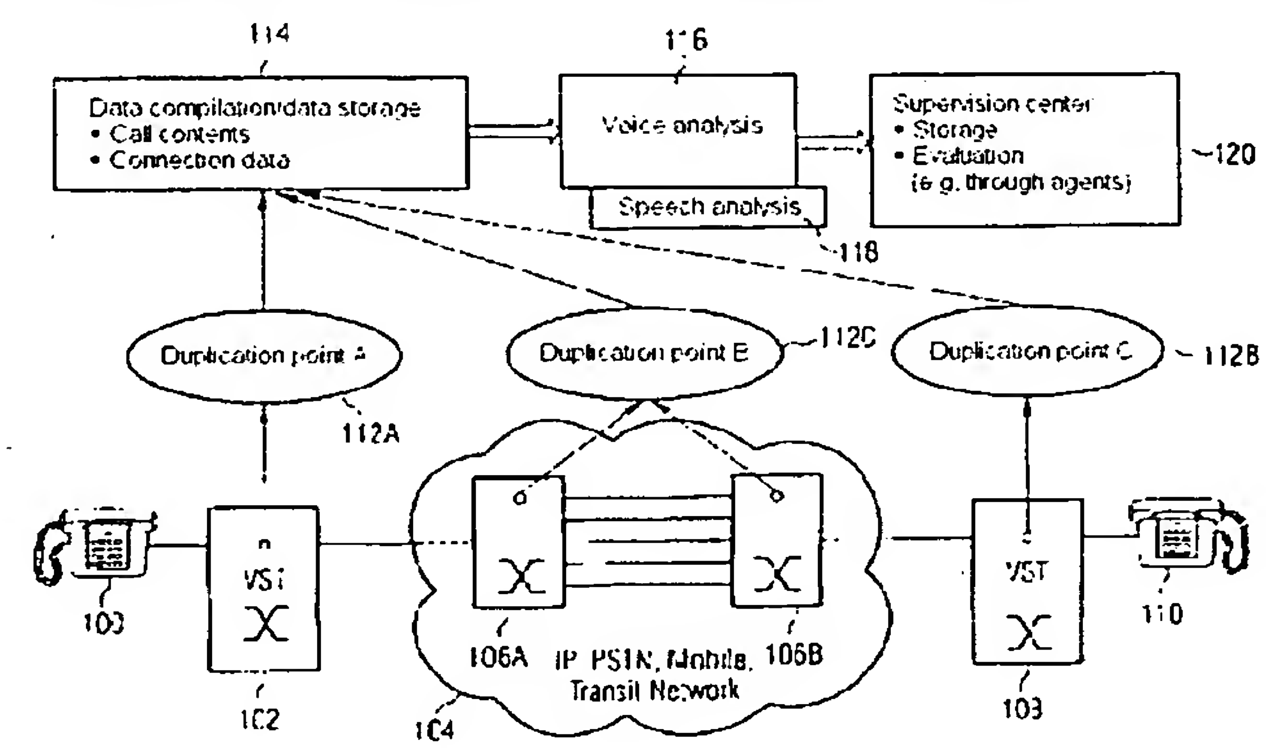
(56) Documents taken into account for judging the possibility of patenting:
US2002/01 50 096 A1
US2001/00 43 687 A1
US2001/00 36 821 A1
US 59 13 161 A
US 64 96 483 B1
US 64 24 701 B1

The following details have been taken from the documents submitted by the registering company

The application for examination is submitted according to § 44 PatG.

(54) Designation: **Procedure, Network Element, and Network Arrangement for the Supervision of Telecommunications**

(57) Summary: The invention applies to the field of legal supervision and compilation of relevant data by the prosecution authorities, generally summarized through the term of Lawful Interception (LI). Well-known LI solutions are always based on endpoint addresses, which lead practically to incomplete LI data and/or to the supervision of a large number of users not subjected to the LI action. According to the invention, a process is planned for the supervision of telecommunications, where a supervision region is selected and supervision parameters are given in the form of speech or voice characteristics. In the supervision region, the entire telecommunications traffic is recorded and submitted to speech and voice analysis, and the portion of the telecommunications traffic matching the supervision parameters is selected for composing supervision data.



Description

[0001] Modern communications networks offer many speech communications possibilities for users through various media. As in the past, much use is made of speech services of classical line telephone networks, often called a fixed network or PSTN. Mobile telephone extensions are also widely used, and with the increase of efficient Internet connections, packet-based speech connections gain an increasing importance.

[0002] The large number of media and protocols of present and future communications networks makes it easier for users with criminal intentions to plan and arrange criminal actions, while at the same time the legal supervision and compilation of relevant data by the prosecuting authorities, generally summarized through the term of Lawful Interception (LI), is made considerably more difficult.

State of Technology

[0003] Until now, LI solutions have always been based on endpoints, a state which requires that for a successful LI action, all the endpoints of a user to be supervised must be known. For instance, the following details must be known: All Internet providers, all fixed network extensions, all mobile endpoints etc. that the user employs. Often, this requirement cannot be fulfilled. For instance, the user can evade LI action through public pay phones or mobile prepaid cards, which are often issued without any identity control or with insufficient identity control, or by employing Internet access that is open to the public, e.g. in Internet cafés or libraries.

[0004] A second problem exists, when an LI action based on endpoints takes up in principle also users, for whom the required LI action is not applicable and for whom a legal supervision order is not applicable, e.g. for family members. This is especially difficult when the endpoints to be supervised include a public extension, because in this case, the LI action inevitably affects uninvolved users.

Nature of the Task

[0005] The invention's purpose is therefore to name a process, a network element, and a network arrangement that avoid the disadvantages mentioned above.

Implementation Example

[0006] This task is implemented by means of a process for the supervision of telecommunications, where the supervision region is selected in the form of speech or voice characteristics. Within the supervision region, the entire telecommunications traffic is recorded and subjected to speech or voice analysis, and the portion of the telecommunications traffic matching the supervision parameters is selected for composing supervision data.

[0007] Supervision regions can be selected (automatically) as geographical or logical regions. In an advantageous form, the selection of a geographical supervision region is done by first recording a roaming profile for a user and then selecting the supervision region according to the limits of the user's roaming field. To this end, a mobile roaming region of the user can be selected. In an additional form, a logical supervision region is defined through at least one address space, e.g. through the address space of a specific Internet provider.

[0008] Speech or voice characteristics of a user or of a group of users can be selected as supervision parameters, where supervision parameters are selected in such a way that each user can be definitely identified.

[0009] Alternatively, instead of supervising individual users, the invention can be employed in a geographical region that contains a higher potential for disturbance, in order to analyze all telecommunications. Speech or voice characteristics are chosen as supervision parameters, which result in recognition of higher aggression potential among all the users operating in the supervision region. In this case user identification is not necessary; the speech or voice evaluation can be based, for instance, on statistical supervision parameters, e.g. sound level, pitch of voice, speech velocity etc.

[0010] In order to fulfill legal requirements and/or diminish the machine computation requirements for speech or voice analysis, telecommunications traffic can be placed in intermediate storage at first and be later evaluated by speech and voice analysis. In this case, instead of placing the telecommunications traffic in intermediate storage, the special case of buffering can also be provided, if the legal requirements demand this. Technical means can be provided in this case to prevent access to buffered, unfiltered, telecommunications traffic, so that no (possibly inadmissible) general supervision takes place in the supervision region.

[0011] The invention foresees also a network element for the supervision of telecommunications, which provides the following:

- Means for receiving telecommunications traffic indicated by a duplication point, where the telecommunications traffic includes connection contents and data referring to the traffic;
- Means for analyzing the connection contents on the basis of supervision parameters containing at least speech or voice characteristics;
- Means for forwarding the portion of the telecommunications traffic matching the supervision parameters to a supervision center.

[0012] In addition, the network element can have storage means for the described buffering of the telecommunications traffic.

[0013] A preferred network element for the supervision of data following a Voice over Internet Protocol contains the following.

- Packet filters for recognizing a Voice over Internet Protocol process;
- Means for intermediate storage of the data stream in the Voice over Internet Protocol process;
- Means for decoding packet-based speech data and for creating a sound stream;

- Means for comparing the sound stream with supervision parameters showing at least speech and voice characteristics;
- Means for conducting the intermediately stored data stream to the supervision center responding to the matching of the sound stream with the supervision parameters.

[0014] Additionally, in the network element prepared according to the invention, there can be means for speech recognition in order to convert the relevant telecommunications into text, as well as means to forward the text to the supervision center.

[0015] The network element prepared according to the invention can be integrated in switching elements in order to reduce the quantities of data by filtering them at the place of their creation.

[0016] The invention concerns also a network arrangement for telecommunications supervision, which includes the following:

- Access switching elements, which supply telecommunications services to terminal devices;
- A connection network;
- One or more duplicating points to duplicate the telecommunications traffic and to forward the duplicated traffic to a supervision network element constructed according to the invention.

[0017] The advantage of the invention is that at first in the supervision region the full telecommunications are taken into account for an LI action, are put to intermediate storage or are buffered, they are then analyzed according to speech or voice characteristics, which when applied makes it possible to implement the destination-oriented and complete supervision of a user or of a certain user group in the supervision region and prevents at the same time the supervision of noninvolved users.

[0018] In another case of application, the invention serves to ascertain an aggression potential. In this action, the evaluation of the aggression potential can take place also in parallel to the first case of application (supervision of selected users).

[0019] The invention can also be employed together with the conventional LI process, where the supervision is based on addresses of the user to be supervised, in order to forward only the telecommunications traffic produced really by the user to be supervised up to a work station in the supervision center and not to the telecommunications traffic created by users to whom the LI action does not apply. With this combination, the inclusion of public endpoints is not complicated. At the same time, the expense for the required speaker recognition is low, because the supervision region is accordingly small (it contains then the endpoints selected for the conventional LI process).

[0020] In the following, an example of implementing the invention is described in detail together with three drawings.

[0021] Fig. 1 shows a network arrangement for the supervision of telecommunications in a schematic diagram.

[0022] Fig. 2 shows schematically the course of the supervision of telecommunications in real time.

[0023] Fig. 3A-B shows schematically the course of the supervision of telecommunications with intermediate storage or buffering of the telecommunications.

[0024] Fig. 1 shows a typical network arrangement with two users, **100** and **110**, given as an example, who are provided through local exchanges, **102** and **108**, with telecommunications services. The two exchanges are connected through a transit network **104**, which has transit exchanges **106A**-B.

[0025] The transit network **104** can be a conventional PSTN transit network, a mobile telephone transit network, a transit network based on the Internet Protocol IP, a transit network based on the Ethernet, or any other transit network. Instead of using the classical local exchanges **102** and **108**, the users **100** and **110** can also be supplied with telecommunications services through Voice over IP servers, mobile radio base stations, and other means of network access, which have not been illustrated.

[0026] Fig. 1 shows also three duplication points **112A**-C, where telecommunications data are duplicated. In this process, both telecommunications contents (e.g. call contents and data) and control information (e.g. protocol-specific signaling information and connection data) are duplicated.

[0027] The duplicated data is forwarded to a first component **114** for data compilation. The first component **114** receives the duplicated data from the duplicating points **112** and controls these points. If there are several duplicating points **112** as in the given example, which duplicate the data referring to the same connection (here referring to the connection between the users **100** and **110**), the first component **114** selects a suitable duplicating point **112** in order to record the data of this connection. Otherwise, the first component **114** receives continuously data from all duplication points **112** and rejects data recorded twice.

[0028] As already explained, the telecommunications traffic can be put at first to intermediate storage by the first component **114**, in order to comply with the legal requirements and/or in order to reduce the computation requirements for means of speech or voice analysis.

[0029] From the first component **114**, the recorded telecommunications data is forwarded to a second component **116** for voice analysis. The voice analysis is carried out with supervision parameters, which represent the voice and speech characteristics of a user to be supervised. These supervision parameters can be obtained from existing speech records containing the speech of the user to be supervised. If there are no speech records of this kind, suitable speech records can be produced by supervising a communications endpoint that can be assigned to the user, who is to be supervised, by means of a classical LI action. In order to prevent the falsification of the supervision parameters by other users of the same endpoint, an LI agent can examine the speech records before they are transformed into supervision parameters.

[0030] The second component **116** can be complemented by a third component **118** for speech recognition (speech-to-text transformation). This is advantageous in combination with the second component **116**, because the analysis of the speech data, which is necessary for the speaker identification provides intermediate results, which can be further used for the speech identification. The speech identification transforms the speech data assigned to the user, who is to be supervised, into machine-readable texts.

[0031] Only the telecommunications data that can be assigned to the user, who is to be supervised, is forwarded to a supervision center **120** (called also Monitoring Center, MC). There, the data is stored and evaluated, e.g. by an LI agent or automatically.

[0032] In this action, the supervision is limited to one supervision region. Various considerations can lead to the specification of a supervision region:

- The specification of the supervision region can take place purely administratively; e.g. the entire (geographical) territory of a country can be specified as a supervision region.
- The supervision region can be restricted to a certain access provider or to his (logical) communications address spaces, if it is known that the user utilizes only this provider. The same can be applied to two or more providers utilized by the user.
- The supervision region can be found automatically by supervising for a certain time an endpoint ready for roaming of the user to be supervised and by determining the supervision region on the basis of the (geographical) roaming region (e.g. as the established roaming region extended by a safety radius). Mobile roaming regions and Internet accesses ready for roaming are especially suitable. In the case of the Internet, the geographical data of IP addresses can be recorded alternatively, from where certain services that can be assigned to the user were called up (e.g. an Email query or Login for Online Banking) in order to create a roaming profile.
- A suitable algorithm can combine the above-mentioned criteria in order to supervise for instance only part of a roaming region situated in the territory of a certain country, because perhaps outside this country, LI measures are subject to other requirements. Likewise, by an "OR" combination of the geographical region with the logical region, a larger supervision region can be defined.

[0033] In Fig. 1, three duplication points **112A-C** are displayed. In practice, only one of these duplication points is necessary. In this process, a duplication point **112A-B** can be arranged in conjunction with a local exchange **102, 108,** or it can be realized centrally in conjunction with the transit network **104** (duplication point **112C**).

[0034] While it is possible to duplicate the entire traffic of the communications network with one duplication point **112C** coupled to the transit network **104**, the requirements of such a duplication point **112C** are quite high. Therefore, it might be advantageous to install duplication points **112A-B** at several peripheral points in the communications network. Depending on the hierarchy and structure of the communications network, the direct subscriber access elements (e.g. local exchanges) or network elements of a higher hierarchical level are offered.

[0035] In both cases, before the speech and/or voice analysis, a pre-selection in the second component **116** is possible on the basis of subscriber addresses (e.g. telephone numbers, IP addresses) carried out for instance by the first component **114** or by the duplication points **112**. The pre-selection can serve to keep up the specified limits of the geographical or logical supervision region. Moreover, the pre-selection through positive lists can basically exclude certain addresses from the supervision and/or the pre-selection through negative lists can basically subject certain addresses to the supervision.

[0036] In this way, an advantageous combination of known LI processes referring to endpoints utilized by the user to be supervised can be achieved with the invention, when for instance the telecommunications traffic produced at the endpoints to be supervised will then be subjected to filtering with automatic speech identification. In other words, the specification of the supervision region can take place through telecommunications addresses with the criteria applicable to current LI actions.

[0037] Various aggregations and integration levels of the elements and components presented in Fig. 1 are possible technically and advantageous; however, these aggregations and integration levels are not legally admitted in all countries. In order to reduce the large data quantities during the supervision of the entire telecommunications traffic already at their origin to the relevant data, it would be desirable to implement a combination of duplication point **112C**, first component **114** and second component **116** directly in connection with one of the transit nodes 106A-B of the transit network **104**. If this is not possible due to legal causes, at least a separate implementation of the above-mentioned combination should be made as near as possible to the data origin, e.g. transit nodes, with a respectively efficient interface.

[0038] A schema of a possible form of the process according to the invention is shown in Fig. 2. In the example presented, the supervision takes place with respect to the connection. At first, a connection set-up **200** is examined to see, whether it is relevant for the LI action (step **202**). In step **202**, the LI relevance is examined with regard to criteria established earlier, e.g. an examination is carried out regarding the question, whether the connection origin or destination belongs to the supervision region.

[0039] If the connection set-up LI is relevant, a query **204** can be provided showing, whether the LI action is carried out on the basis of DN subscriber numbers (DN = Directory Number). If this occurs [point (a) of the process], the agreement of the subscriber numbers (of the calling and called subscriber) with the LI criteria is examined (step **206**); this is branched in step **208** to point (1) of the process in case of non-agreement.

[0040] If the query in step **202** is negative, i.e. if the connection set-up is not LI-relevant, a decision is made in step **222** according to specifications, whether for the present supervision region, the aggression potential of the connected subscribers should be examined. If the query in step **222** is negative [point (1) of the process], step **224** is examined, showing whether the data comes from a delayed data analysis (post-processing, see below with reference to **Fig. 3**) If yes, the post-processing is finished. If no, the LI connection is finished and the subscriber connection can be continued without LI influence.

[0041] If however, query **208** or **222** is fulfilled, or if query **204** is not fulfilled [point (b) of the process], the speech or voice analysis (step **210**) is reached. If the detected parameters, which characterize the current speaker, agree with the supervision parameters at least for one speaker (step **212**), then in step **214** [point (2) of the process] the forwarding of telecommunications data to the supervision center MC (step **216**) is initiated; all the data and contents related to this connection are registered after a positive identification of a user to be supervised without any further examination. In addition, the telecommunications traffic transferred until the secured positive identification can be put to intermediate storage - this is not displayed. In case of a positive identification, this section of a call preceding the identification is forwarded also to the supervision center; otherwise, it is rejected.

[0042] If in step **214**, no agreement could be established between one of the speakers and one of the users to be supervised, a disposition analysis takes places with regard to the aggression potential. If aggressions are recognized, e.g. by means of keywords in the evaluation of a speech recognition (element **118** from Fig. 1) and/or according to the tone position, speech velocity, and similar factors, additional measures can be initiated. In the given example, branching is carried out to point (2) of the process in order to forward the data to the supervision center MC (step **216**). In the supervision center, with the amount of connections established in a certain region with a higher aggression potential, a rising disturbance for instance can be found, and further measures can be initiated, e.g. an intensified visual observation in the region or a strengthening of the security forces.

[0043] However, if no aggressions can be derived from the disposition analysis in steps **218** and **220**, branching can be carried out to point (1) of the process (see above). An alternative way is to return to point (b) of the process in order to examine, whether a speaker to be supervised has joined the connection, e.g. when the connection is enlarged to a conference or when a user actually to be supervised asks at first unsuspicious users to establish the connection in order to join this connection later on.

[0044] With regard to Fig. 3A-B, the course will be described below for the process according to the invention, if intermediate storage of the telecommunications must be provided. As mentioned already, both technical and legal backgrounds can be decisive for an intermediate storage of the telecommunications traffic followed by the evaluation in post-processing. In these cases, the real time process given in Fig. 2 is not directly applicable.

[0045] Fig. 3A shows the course for the storage of the telecommunications traffic. A connection set up in step **300** is duplicated in all data, both call data CC (CC = Call Content) and supervision relevant data IRI (IRI = Interception Related Information; e.g. date of call, duration, etc.), through duplication points **112** (step **302**) and is put to intermediate storage (step **304**), before the connection is terminated regularly in step **306**.

[0046] Fig. 3B shows the post-processing which can take place subsequently. Due to the real time requirements, which do not exist in this case, the post-processing can for instance be carried out also during times of lower traffic. For the post-processing, data are taken in step **310** from the intermediate storage, and if an analysis of this data is foreseen (query **312**), it will be checked in step **314**, whether the LI action takes place based on the subscriber addresses DN. If yes, the processing will be continued in point (a) of the process from Fig. 2. If no, it will be checked in step **316**, whether a speech and/or voice analysis should take place. If yes, the processing will be continued point (b) of the process from Fig. 2; otherwise, the post-processing is terminated.

[0047] A special case of voice analysis for the purpose of speaker recognition results, if any protocol of Voice over IP (VoIP) is employed, e.g. H.323, SIP, or proprietary derivates such as Net2Phone with the respective signaling and/or control protocols. In this matter, a distinction is made between two case: The voice analysis with the purpose of speaker identification takes place directly in the duplication point, which is also called a trial, or in a downstream analysis component, as shown for instance in Fig. 1.

[0048] For speaker identification in the duplication point, a filter is used to recognize VoIP traffic, e.g. a Berkeley Packet Filter (BPF), which works for instance with IP addresses or TCP/UDP port numbers. The VoIP traffic is decomposed into signaling information and useful information (payload), e.g. H.225 (signaling channel), H.245 (control channel), and RTP (RTP = Real Time Protocol, which forwards the payload). The signaling information or control information is often called meta-data.

[0049] The supervision parameters characterizing the user/users to be supervised are stored in the duplication point or can be retrieved through the duplication point. In the duplication point, the control and/or signaling information is decoded and the payload stream is extracted. The Codec used for speech coding is determined. If necessary, the payload stream and the associated control and/or signaling information is buffered, and a temporary sorting of the packets of the payload stream takes place.

[0050] As it is not possible, in general, to produce directly the speech parameters from the VoIP-coded speech data for comparison with the supervision parameters, using the determined Codec the VoIP-coded speech data is initially transformed into spoken speech, into a sound stream, or into an intermediate format. The sound stream is then analyzed like spoken speech (see above), and when a user to be supervised is positively identified, the telecommunications traffic to be assigned to this VoIP stream including the data buffered until then is forwarded to the supervision center.

[0051] However, if the analysis of the VoIP traffic is downstream, possibly as displayed in Fig. 1, the above-mentioned functions are implemented in a suitable component **116**. In this case, the duplication point serves only to filter/identify and forward the VoIP traffic to this component for voice/speaker analysis.

[0052] Naturally, it is also possible in connection with VoIP traffic and even necessary under the circumstances already mentioned to put at first the VoIP traffic to intermediate storage and then analyze it for speaker identification.

[0053] The filtering in the duplication point can also include for VoIP traffic a filtering according to preset source and/or destination addresses in order to restrict the LI action to a geographical or logical region or to certain addresses employed generally by the user.

[0054] As shown already, the invention can be used on the one hand in the framework of known LI actions that are implemented on the basis of known communications addresses of users to be supervised, and can be completed for the purpose of identifying among the possible users of a communications endpoint those users who really are to be supervised and to forward only their communications to a supervision center, and on the other hand in order to analyze the entire communications traffic in a region that can be easily defined and to forward only calls of users to be supervised to a supervision center.

[0055] In the given examples of implementation, the invention was described as an example only with regard to fixed network communications and VoIP communications. Basically, the invention is suitable for supporting the supervision of telecommunications of all kinds. Therefore, the invention is not restricted to the implementation example but includes all forms of telecommunications, e.g. speech connections between two subscribers, conferences with any number of subscribers, connections to announcement systems or

answering equipment over any telecommunications systems and protocols.

Patent Claims

1. Process for the supervision of telecommunications including the following steps:

- Selection of a supervision region;
- Specification of supervision parameters in the form of speech or voice characteristics;
- Compilation of the entire telecommunications traffic in the supervision region;
- Implementation of a speech and voice analysis of the recorded telecommunications traffic; and
- Selection of the portion of the telecommunications traffic matching the supervision parameters for composing supervision data.

2. Process according to Claim 1, by which as supervision region, a geographical supervision region is selected, where at first a roaming profile is recorded for a user and the supervision region is selected on the basis of the limits of the roaming region of the user.

3. Process according to Claim 2, by which a geographical supervision region is selected on the basis of the limits of a mobile roaming region of the user.

4. Process according to Claim 1, which as supervision region defines a logical supervision region defined by means of at least one address space.

5. Process according to one of Claims 1 to 4, which as supervision parameters selects the speech or voice characteristics of a user or of a group of users, where supervision parameters are selected in such a way that each user can be definitely identified.

6. Process according to one of Claims 1 to 3, which as supervision region selects a geographical region selected, that contains a higher disturbance potential and speech or voice characteristics are selected as supervision parameters allows the determination of a higher aggression potential among all users operating in the supervision region.

7. Process according to one of Claims 1 to 6, by which after the step for compiling the telecommunications traffic, a step for intermediate storage of the telecommunications traffic follows and by which the speech and voice analysis is implemented downstream for the telecommunications traffic put to intermediate storage.

8. Network element (**116**) for telecommunications supervision including the following:

- Means for receiving telecommunications traffic indicated by a duplication point (**112**), where the telecommunications traffic includes connection contents and data referring to the connection;
- Means for analyzing the connection contents on the basis of supervision parameters containing at least speech and voice characteristics;
- Means for forwarding the portion of the telecommunications traffic matching the supervision parameters to a supervision center (**120**).

9. Network element (**116**) according to Claim 8, which includes also means of storage (**114**) for the intermediate storage of the telecommunications traffic.

10. Network element (**116**) according to Claim 8 for the supervision of data according to a process of Voice over Internet Protocol, where the network (**116**) includes the following:

- Packet filters for recognizing a process of the Voice over Internet Protocol;
- Means for intermediate storage of the data stream in the process of the Voice over Internet Protocol;
- Means for decoding packet based speech data and for creating a sound stream;
- Means for comparing the sound stream with supervision parameters showing at least speech and voice characteristics;
- Means for forwarding the intermediately stored data stream to the supervision center responding to matching of the sound stream with the supervision parameters

11. Network element (**116**) according to one of Claims 8 to 10, which includes also means for speech recognition (**118**) as well as means for forwarding the portion converted into text of the telecommunications traffic matching the supervision parameters to the supervision center (**120**).

12. Switching element (**102, 106, 108**) of a telecommunications network, in which a network element (**116**) is integrated according to one of Claims 8 to 11.

13. Network arrangement for telecommunications supervision including the following:

- Access switching elements (**102, 108**), which supply telecommunications services to terminal devices (**100, 110**);
- A connection network (**104**);
- One or several duplicating points (**112**) in order to duplicate the telecommunications traffic and to forward the duplicated traffic to supervision network element (**116**) according to one of Claims 8 to 11.

There follow drawings on 3 pages

Attached drawings

FIG 1

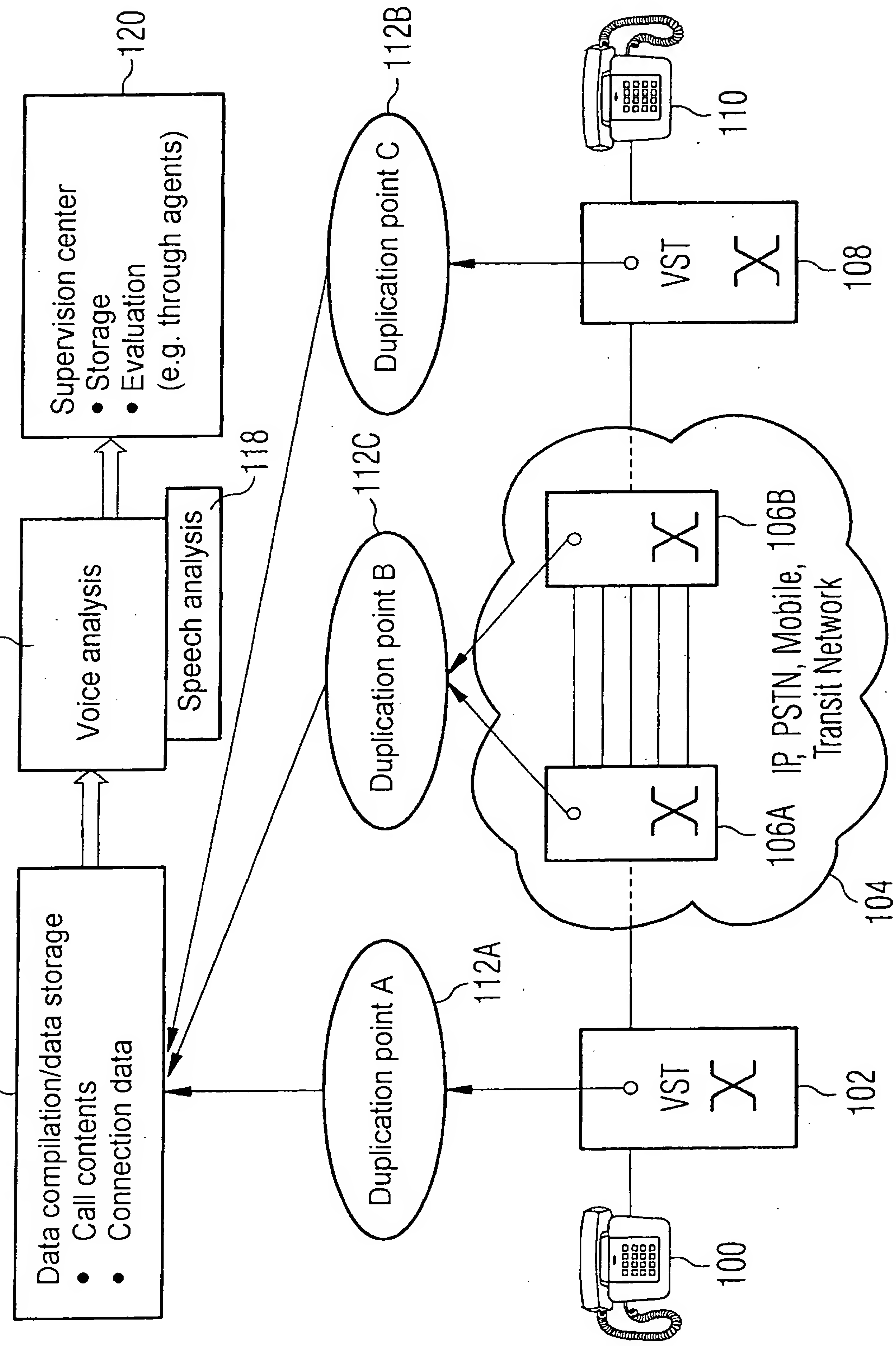


FIG 2

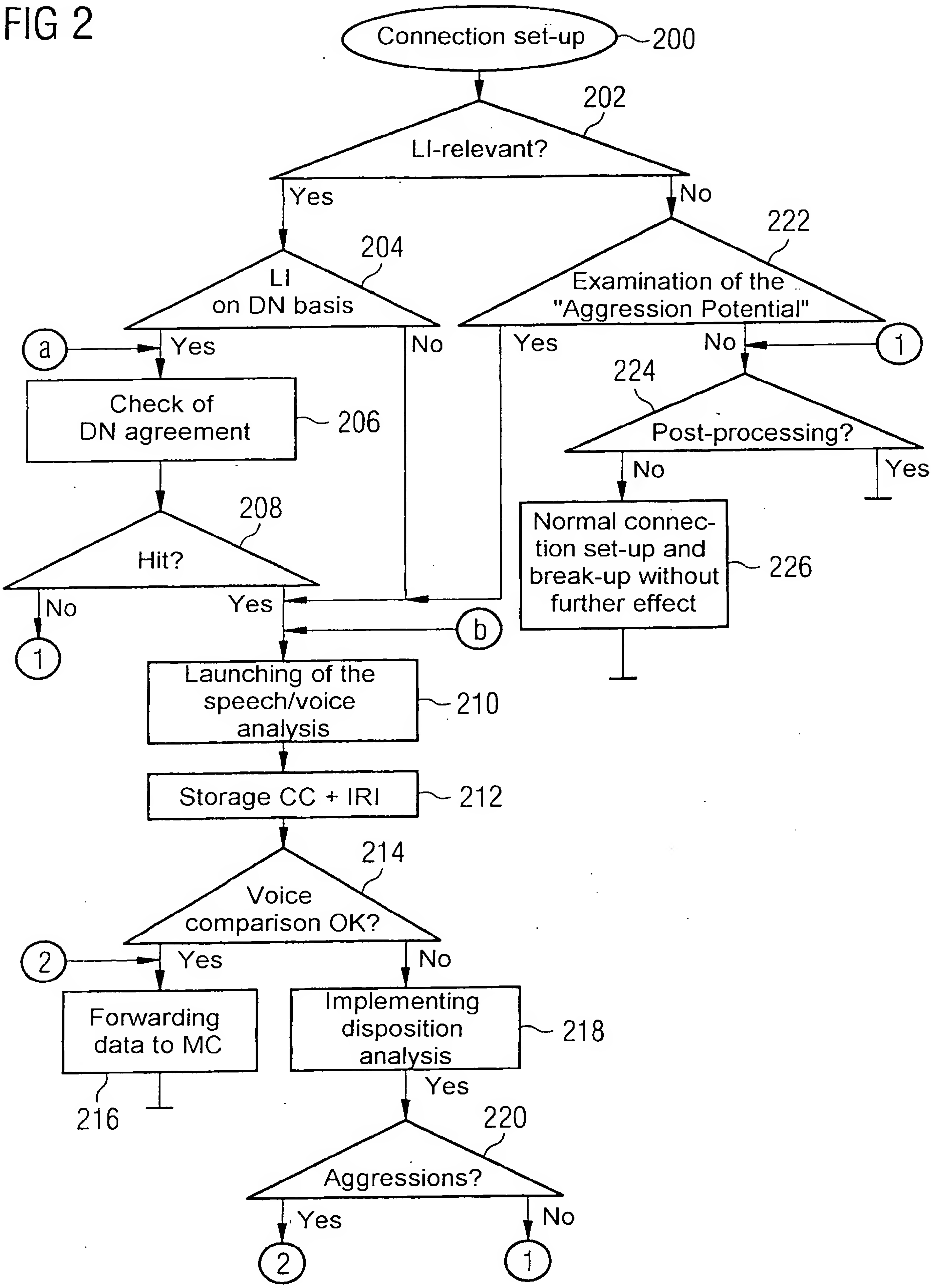

Connection set-up
200
202
LI-relevant?
Yes
No
222
204
LI
on DN basis
Examination of the
"Aggression Potential"
a
Yes
No
Yes
No
1
224
Check of
DN agreement
206
Post-processing?
No
Yes
208
Normal connec-
tion set-up and
break-up without
further effect
226
Hit?
No
Yes
b
1
Launching of the
speech/voice
analysis
210
Storage CC + IRI
212
214
Voice
comparison OK?
2
Yes
No
Forwarding
data to MC
Implementing
disposition
analysis
218
216
Yes
220
Aggressions?
Yes
No
2
1

FIG 3A

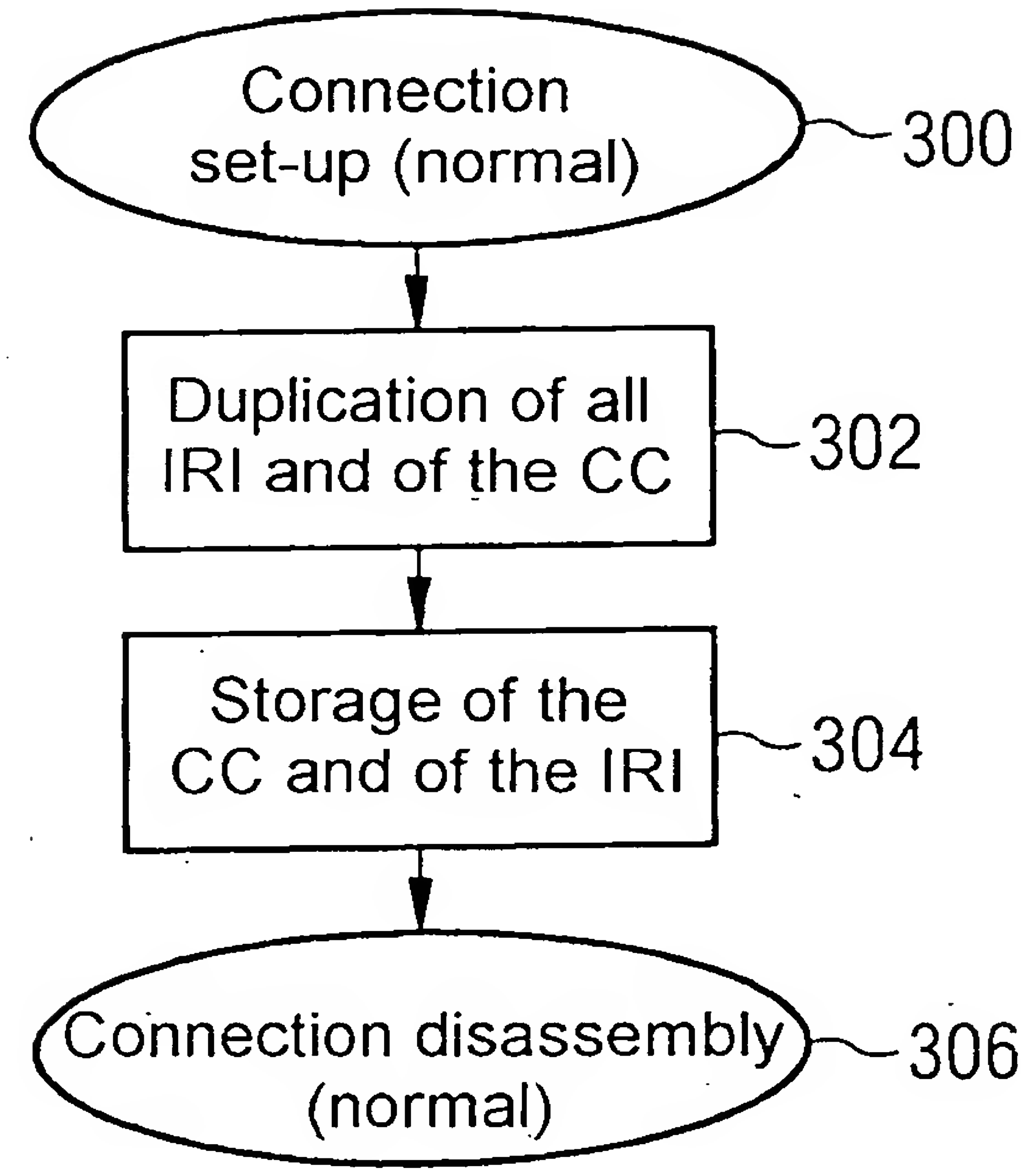
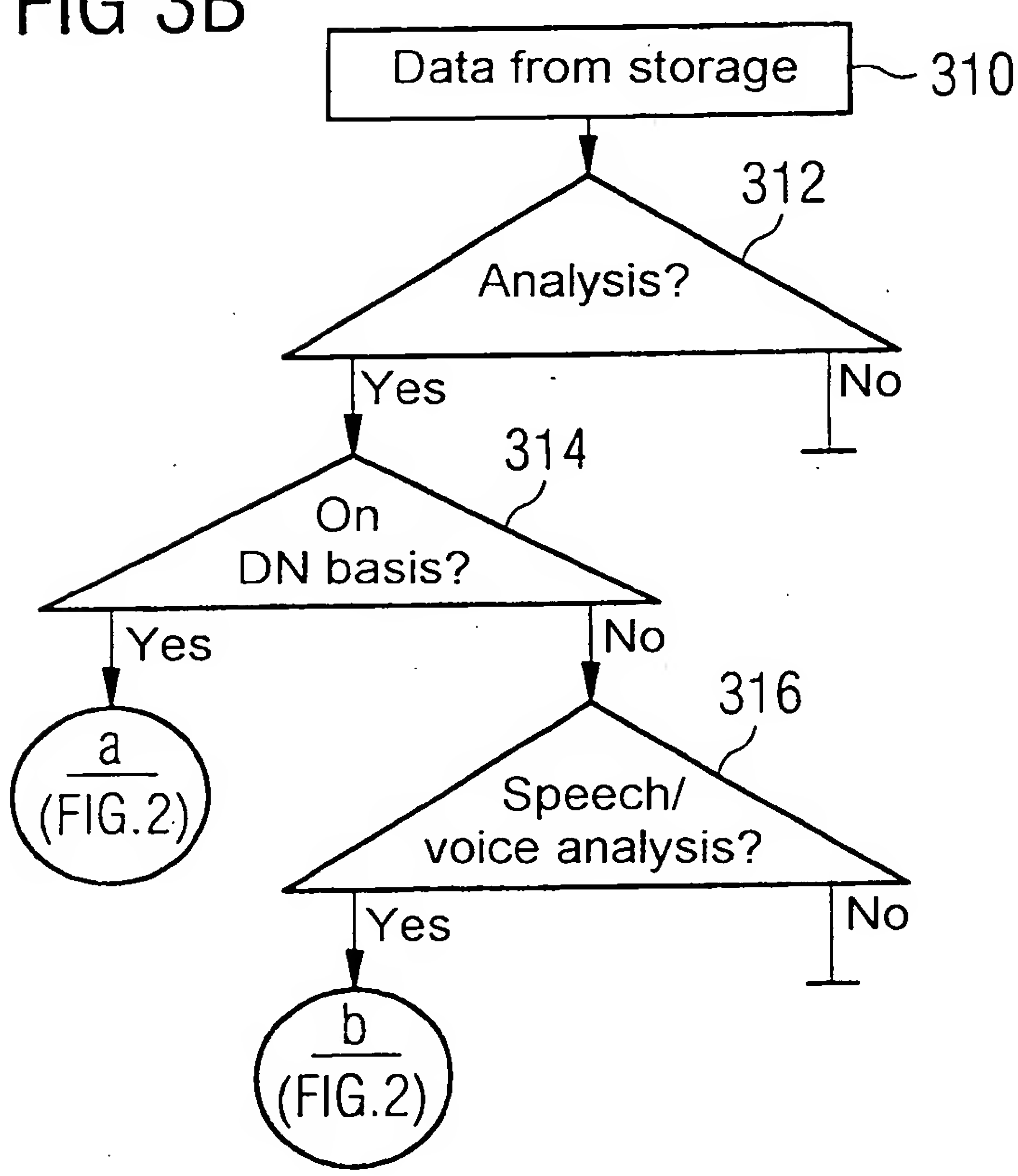

Connection set-up (normal)
300
Duplication of all IRI and of the CC
302
Storage of the CC and of the IRI
304
Connection disassembly (normal)
306

FIG 3B


Data from storage
310
312
Analysis?
Yes
No
314
On DN basis?
Yes
No
a (FIG.2)
316
Speech/ voice analysis?
Yes
No
b (FIG.2)